

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

| | Type of Search | Vendors and cost where applicable |
|---|--------------------------|-----------------------------------|
| Searcher: <u>Q. V. Jones - J. J. J.</u> | NA Sequence (#) _____ | STN _____ |
| Searcher Phone #: <u>308-4501</u> | AA Sequence (#) <u>1</u> | Dialog _____ |
| Searcher Location: <u>Biotech Lib.</u> | Structure (#) _____ | Questel/Orbit _____ |
| Date Searcher Picked Up: <u>7/3/01</u> | Bibliographic _____ | Dr. Link _____ |
| Date Completed: <u>7/3/01</u> | Litigation _____ | Lexis/Nexis _____ |
| Searcher Prep & Review Time: _____ | Fulltext _____ | Sequence Systems <u>ABSS02</u> |
| Clerical Prep Time: <u>3 min</u> | Patent Family _____ | WWW/Internet _____ |
| Online Time: <u>2 min</u> | Other _____ | Other (specify) _____ |

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using BL model

Run on: July 3, 2001, 16:13:43 ; Search time 23 Seconds
(without alignments)
1241.474 Million cell updates/sec

Title: US-09-454-223-2
Perfect score: 2487
Sequence: 1 MLPLSLMLVLVQPLGNLGA.....TDPQVSHGTGFTSRGLKL 471

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_0601:*

- 1: /SIDS8/gcgdata/geneseq/AA1980.DAT:*
- 2: /SIDS8/gcgdata/geneseq/AA1981.DAT:*
- 3: /SIDS8/gcgdata/geneseq/AA1982.DAT:*
- 4: /SIDS8/gcgdata/geneseq/AA1983.DAT:*
- 5: /SIDS8/gcgdata/geneseq/AA1984.DAT:*
- 6: /SIDS8/gcgdata/geneseq/AA1985.DAT:*
- 7: /SIDS8/gcgdata/geneseq/AA1986.DAT:*
- 8: /SIDS8/gcgdata/geneseq/AA1987.DAT:*
- 9: /SIDS8/gcgdata/geneseq/AA1988.DAT:*
- 10: /SIDS8/gcgdata/geneseq/AA1989.DAT:*
- 11: /SIDS8/gcgdata/geneseq/AA1990.DAT:*
- 12: /SIDS8/gcgdata/geneseq/AA1991.DAT:*
- 13: /SIDS8/gcgdata/geneseq/AA1992.DAT:*
- 14: /SIDS8/gcgdata/geneseq/AA1993.DAT:*
- 15: /SIDS8/gcgdata/geneseq/AA1994.DAT:*
- 16: /SIDS8/gcgdata/geneseq/AA1995.DAT:*
- 17: /SIDS8/gcgdata/geneseq/AA1996.DAT:*
- 18: /SIDS8/gcgdata/geneseq/AA1997.DAT:*
- 19: /SIDS8/gcgdata/geneseq/AA1998.DAT:*
- 20: /SIDS8/gcgdata/geneseq/AA1999.DAT:*
- 21: /SIDS8/gcgdata/geneseq/AA2000.DAT:*
- 22: /SIDS8/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------|--------------------|
| 1 | 1110 | 44.6 | 261 | AA36701 | CD40-L. Homo sapi |
| 2 | 1110 | 44.6 | 261 | AA35969 | Human CD40-L type |
| 3 | 1110 | 44.6 | 261 | AA57469 | CD40 ligand. Synt |
| 4 | 1110 | 44.6 | 261 | AA85486 | Human CD40 ligand. |
| 5 | 1110 | 44.6 | 261 | AA34659 | Amino acid sequenc |
| 6 | 1110 | 44.6 | 261 | AA71751 | Human CD40 ligand. |
| 7 | 1110 | 44.6 | 261 | AA41178 | CD40 ligand. Homo |
| 8 | 1110 | 44.6 | 261 | AA39988 | Human CD40-L prote |
| 9 | 1110 | 44.6 | 261 | AA25155 | Human soluble CD40 |
| 10 | 1110 | 44.6 | 261 | AA96993 | Human CD40 ligand. |
| 11 | 1110 | 44.6 | 261 | AA67612 | Amino acid sequenc |

| | | | | | |
|----|--------|------|-----|---------|---------------------|
| 12 | 1110 | 44.6 | 261 | AA37806 | Human CD40 ligand |
| 13 | 1110 | 44.6 | 261 | AA09122 | CD40 ligand/zipper |
| 14 | 1105 | 44.4 | 261 | AA09119 | CD40 ligand/zipper |
| 15 | 1104 | 44.4 | 261 | AA09121 | CD40 ligand/zipper |
| 16 | 1102 | 44.3 | 261 | AA49548 | Predicted sequence |
| 17 | 1101 | 44.3 | 261 | AA09120 | CD40 ligand/zipper |
| 18 | 1100 | 44.2 | 261 | AA09114 | Human CD40L mutle |
| 19 | 1099 | 44.2 | 261 | AA76125 | Human CD40L. Hom |
| 20 | 1099 | 44.2 | 261 | AA09113 | Human CD40L mutle |
| 21 | 1099 | 44.2 | 261 | AA09129 | CD40 ligand/zipper |
| 22 | 1098 | 44.1 | 261 | AA09115 | Human CD40L mutle |
| 23 | 1094.5 | 44.0 | 273 | AA09118 | Yeast GCN4 leucine |
| 24 | 1094.5 | 44.0 | 273 | AA09116 | Human trimeric CD4 |
| 25 | 1094.5 | 44.0 | 273 | AA39940 | Human CD40-L/FC fu |
| 26 | 1089 | 43.8 | 473 | AA36703 | Human CD40-L/FC fu |
| 27 | 1089 | 43.8 | 473 | AA39939 | Human CD40-L/FC fu |
| 28 | 1086 | 43.7 | 294 | AA09123 | CD40 ligand/zipper |
| 29 | 1064 | 42.8 | 375 | AA77990 | Human SP-D amino a |
| 30 | 939.5 | 37.8 | 260 | AA58217 | Feline mature CD15 |
| 31 | 921.5 | 37.1 | 211 | AA58218 | Feline mature CD15 |
| 32 | 907.5 | 36.5 | 260 | AA58215 | Canine mature CD15 |
| 33 | 889.5 | 35.8 | 211 | AA58216 | Canine mature CD15 |
| 34 | 879 | 35.3 | 371 | AA75642 | Bovine conglutinin |
| 35 | 831 | 33.4 | 260 | AA53970 | Mouse CD40-L type |
| 36 | 831 | 33.4 | 260 | AA41179 | CD40 ligand. Mus |
| 37 | 831 | 33.4 | 260 | AA39937 | Mouse CD40-L prote |
| 38 | 829.5 | 33.4 | 351 | AA18780 | Bovine conglutinin |
| 39 | 829.5 | 33.4 | 351 | AA13672 | Bovine conglutinin |
| 40 | 827 | 33.3 | 260 | AA71750 | Murine CD40-L. Mu |
| 41 | 825 | 33.2 | 260 | AA36702 | Murine CD40-L. Mu |
| 42 | 812 | 32.6 | 280 | AA39941 | Mouse trimeric CD4 |
| 43 | 807 | 32.4 | 177 | AA10874 | CD40L/TF-beta hybr |
| 44 | 805 | 32.4 | 177 | AA10873 | CD40L/TF-alpha hybr |
| 45 | 769 | 30.9 | 182 | AA09126 | CD40 ligand/zipper |

ALIGNMENTS

RESULT 1

| | | |
|----|---|------------------------------|
| ID | AA36701 | standard; Protein; 261 AA. |
| AC | AA36701 | |
| DT | 12-AUG-1993 | (first entry) |
| DE | CD40-L. | |
| DE | Human: CD40-L; CD40: type II; membrane; polypeptide; extracellular; | |
| KW | transmembrane; induction; intracellular; soluble; activity; B cell; | |
| KW | proliferation; induction; antibody; secretion; i9e; agonist; | |
| KW | antagonist; binding assay. | |
| OS | Homo sapiens. | |
| FT | Key | Location/Qualifiers |
| FT | Region | 47..261 |
| FT | | /note="Extracellular region" |
| PN | WO9308207-A. | |
| PD | 29-APR-1993. | |
| PF | 23-OCT-1992; | 92WO-US08990. |
| PR | 25-OCT-1991; | 91US-0783707. |
| PR | 05-DEC-1991; | 91US-0805723. |
| PA | (IMMUNEX CORP. | |
| PI | Armstrong RJ, Fanslow WC, Spriggs MK; | |

DR WPI, 1993-152417/18.
N-PSDB; AAQ41506.

XX New cytokine CD40-L as CD40 agonist and antagonist - is used for
PT treating allergies, lupus, rheumatoid arthritis,
XX graft-versus-host disease and insulin-dependent diabetes mellitus
PS Claim 1: Fig 2; 80pp; English.

XX This sequence represents human CD40-L polypeptide which binds to CD40.
CC CD40-L is a type II membrane polypeptide which has an extracellular
CC region at its C-terminus, a transmembrane region and an
CC intracellular region at its N-terminus. A soluble form of CD40-L
CC lacks the transmembrane domain. CD40-L activity is mediated by
CC binding with CD40 an d induces B cell proliferation and induction of
CC antibody secretion, including IgE. Membrane bound CD40-L acts as a
CC CD40 agonist and soluble CD40-L acts as a CD40 antagonist. CD40-L
XX can be used in a binding assay to detect cells expressing CD40.
SQ Sequence 261 AA;

Query Match 44.6%; Score 1110; DB 14; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.3e-73;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 HRRDKIEDENLHEDYFMKTIQRCNTGERSLSLNCCEIKSQEGFVNDIMLNKEETK 316
DB 47 HRRDKIEDENLHEDYFMKTIQRCNTGERSLSLNCCEIKSQEGFVNDIMLNKEETK 106
OY 317 KENSEFMOKGQONPOIAAHYISEASSKTSVLOMAEKGYTSMNVLTLNGKQLTVKQ 376
DB 107 KENSEFMOKGQONPOIAAHYISEASSKTSVLOMAEKGYTSMNVLTLNGKQLTVKQ 166
OY 377 GLYTYIAOVTFCSNREASSQAPFIASICLKSPGRFERILLRAANTHSSAKPCGQOSIHG 436
DB 167 GLYTYIAOVTFCSNREASSQAPFIASICLKSPGRFERILLRAANTHSSAKPCGQOSIHG 226
OY 437 GFELQPGASVFNVTDPQSVSHGTGFTSFGILKL 471
DB 227 GFELQPGASVFNVTDPQSVSHGTGFTSFGILKL 261

RESULT 2
ID AAR53969 standard; Protein; 261 AA.
XX
AC AAR53969;
XX
DT 11-JAN-1995 (first entry)

XX Human CD40-L type II transmembrane protein.
XX
KW Leucine zipper; trimerisation; trimeric CD0-L; fusion protein;
KW hetero-oligomer; homo-oligomer; type II transmembrane protein;
XX soluble CD40-L; tumour necrosis factor family.
OS
XX Homo sapiens.

XX Key Location/Qualifiers
XX Region 50..261
XX /label= "extracellular region
XX /note= "soluble CD40-L"

XX MO9410308-A.
XX PD 11-MAY-1994.
XX PF 20-OCT-1993; 93MO-US10034.
XX PR 23-OCT-1992; 92US-0969703.
XX PR 13-AUG-1993; 93US-0107353.

PA (IMMV) IMMUNEX CORP.

XX Springs MK, Srinivasan S;
XX WPI, 1994-167465/20.
DR N-PSDB; AAR53969.

XX Prepn. of soluble oligomeric mammalian proteins - using host
PT cells to express a fusion protein comprising a leucine zipper
PT domain and a heterologous mammalian protein
XX
PS Example 1; Page 24; 35pp; English.

XX A DNA fragment encoding the extracellular (soluble) region of human
CC CD40-L was ligated to a synthetic oligonucleotide sequence coding
CC for a leader peptide, a 33 amino acid leucine zipper sequence
CC (AAR53968) and the flag (RTM) linker sequence. Cells expressing the
CC fusion construct are grown to accumulate oligomeric, soluble CD40-L
CC in the supernatant. The leucine zipper sequence spontaneously
CC trimerises in solution and fusion proteins comprising
CC the sequence fused to a heterologous mammalian protein also form
CC oligomers.

Query Match 44.6%; Score 1110; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.3e-73;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 HRRDKIEDENLHEDYFMKTIQRCNTGERSLSLNCCEIKSQEGFVNDIMLNKEETK 316
DB 47 HRRDKIEDENLHEDYFMKTIQRCNTGERSLSLNCCEIKSQEGFVNDIMLNKEETK 106
OY 317 KENSEFMOKGQONPOIAAHYISEASSKTSVLOMAEKGYTSMNVLTLNGKQLTVKQ 376
DB 107 KENSEFMOKGQONPOIAAHYISEASSKTSVLOMAEKGYTSMNVLTLNGKQLTVKQ 166
OY 377 GLYTYIAOVTFCSNREASSQAPFIASICLKSPGRFERILLRAANTHSSAKPCGQOSIHG 436
DB 167 GLYTYIAOVTFCSNREASSQAPFIASICLKSPGRFERILLRAANTHSSAKPCGQOSIHG 226
OY 437 GFELQPGASVFNVTDPQSVSHGTGFTSFGILKL 471
DB 227 GFELQPGASVFNVTDPQSVSHGTGFTSFGILKL 261

RESULT 3
ID AAR57469 standard; Protein; 261 AA.
XX
AC AAR57469;
XX
DT 23-MAR-1995 (first entry)

XX CD40 ligand.
XX
KW Probe; primer; PCR; amplify; polymerase chain reaction; detection;
KW mutation; CD40 ligand gene; IGM; ss.

XX Synthetic.
XX
XX MO9417196-A.
XX PN 04-AUG-1994.
XX PD 21-JAN-1994; 94MO-US00786.
XX PF 22-JAN-1993; 93US-0009258.
XX PR 20-JAN-1994; 94US-0184422.

XX (IMMV) IMMUNEX CORP.

PI Armlage RJ, Davison BL, Fanslow WC, Renshaw BR;
 PI Spriggs MK, Widmer MB;
 XX WPI; 1994-264109/32.
 DR N-PSDB; AAO67123.
 XX
 PT Method for detecting mutation in CD 40 ligand gene - comprises
 PT amplification of nucleic acid, and mutational analysis
 XX
 PS Disclosure: Page 22-24; 38pp; English.
 XX
 CC This sequence is encoded by the CD40 ligand gene. Mutations within
 CC the CD40 ligand gene were identified by the method of the invention.
 CC The method comprises isolating DNA from an individual and selectively
 CC amplifying the isolated DNA derived from the CD40 ligand gene. The
 CC amplification product is then analysed to determine if there is a
 CC mutation present and determining if a protein expressed from the
 CC ligand gene will bind CD40. The detection of mutations in the CD40
 CC ligand gene allows subsequent treatment of a syndrome resulting in
 CC elevated levels of serum IgM and diminished levels of other Ig
 CC isotypes, due to mutation in the CD40 ligand gene. ie. X-linked
 CC hyperIgM syndrome.
 CC
 XX Sequence 261 AA;
 SO
 Query Match 44.6%; Score 1110; DB 15; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.3e-73;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 257 HRRLDKIEDERNLHEDVEFKTIORCNTGERSLSLNCCEIKSOFEGFYVDIMLNKEETK 316
 DB 47 HRRDKIEDERNLHEDVEFKTIORCNTGERSLSLNCCEIKSOFEGFYVDIMLNKEETK 106
 OY 317 KENSFEMQKDONPOIAHVISSEASKTTSVLOWAEKGYTMSNNLVTLENGKQLTVKRG 376
 DB 107 KENSFEMQKDONPOIAHVISSEASKTTSVLOWAEKGYTMSNNLVTLENGKQLTVKRG 166
 OY 377 GLYIIYAQVTFCSNREASSQAPFIASICLKSPGFERILLRAANTHSSANPCGQOSIHLG 436
 DB 167 GLYIIYAQVTFCSNREASSQAPFIASICLKSPGFERILLRAANTHSSANPCGQOSIHLG 226
 OY 437 GFVELQPGASVFNVTDPQSVSHCTGFTSGFLKL 471
 DB 227 GFVELQPGASVFNVTDPQSVSHCTGFTSGFLKL 261
 RESULT 4
 AAR85486
 ID AAR85486 standard; Protein; 261 AA.
 XX
 AC AAR85486;
 XX
 DT 18-MAR-1996 (first entry)
 XX
 DE Human CD40 ligand.
 XX
 KW High density membrane-bound CD40 ligand; B-lymphocyte; B-cell;
 KW differentiation; proliferation; baculovirus; Spodoplera frugiperda;
 KW Sf9; insect cell culture; tumour necrosis factor receptor.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Modified-site 6..8
 XX Domain 23..46
 XX Modified-site 239..241
 XX /label= Transmembrane_domain
 XX /label= N-glycosylation_site
 XX
 PN W09529335-A1.

PD 09-NOV-1995.
 XX
 PF 28-APR-1995; 95WO-US05448.
 XX
 PR 28-APR-1994; 94US-0234580.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM PHARM INC.
 XX
 PI Castle BE, Kehry M;
 XX
 DR WPI; 1995-393038/50.
 DR N-PSDB; AAT05763.
 XX
 PT High density membrane bound CD40 ligand - for stimulating the
 PT proliferation of B cells in vitro or in vivo, partic. for producing
 PT differentiated cells
 XX
 PS Disclosure: Fig 1; 74pp; English.
 XX
 CC Human high-density, membrane-bound (hmbd) CD40 ligand (AAR85486) is a
 CC type 2 membrane glycoprotein expressed in activated T-cells. It is
 CC a member of the tumour necrosis factor receptor family. hmbdCD40
 CC induces long-term proliferation of B-cells in culture. These
 CC proliferating B-cells can be induced to differentiate into antibody-
 CC prodg. cells. Proliferation and differentiation in the presence of an
 CC antigen preferentially proliferates or selects differentiating B-cells
 CC to produce antibodies specific to that antigen. Recombinant hmbdCD40
 CC is obtd. by incorporating encoding DNA (AAT05763) into a baculovirus
 CC vector that is then used to transfect sf9 insect cells.
 XX
 SO Sequence 261 AA;
 Query Match 44.6%; Score 1110; DB 16; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.3e-73;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 257 HRRLDKIEDERNLHEDVEFKTIORCNTGERSLSLNCCEIKSOFEGFYVDIMLNKEETK 316
 DB 47 HRRDKIEDERNLHEDVEFKTIORCNTGERSLSLNCCEIKSOFEGFYVDIMLNKEETK 106
 OY 317 KENSFEMQKDONPOIAHVISSEASKTTSVLOWAEKGYTMSNNLVTLENGKQLTVKRG 376
 DB 107 KENSFEMQKDONPOIAHVISSEASKTTSVLOWAEKGYTMSNNLVTLENGKQLTVKRG 166
 OY 377 GLYIIYAQVTFCSNREASSQAPFIASICLKSPGFERILLRAANTHSSANPCGQOSIHLG 436
 DB 167 GLYIIYAQVTFCSNREASSQAPFIASICLKSPGFERILLRAANTHSSANPCGQOSIHLG 226
 OY 437 GFVELQPGASVFNVTDPQSVSHCTGFTSGFLKL 471
 DB 227 GFVELQPGASVFNVTDPQSVSHCTGFTSGFLKL 261
 RESULT 5
 AAW34669
 ID AAW34669 standard; Protein; 261 AA.
 XX
 AC AAW34669;
 XX
 DT 16-FEB-1998 (first entry)
 XX
 DE Amino acid sequence for CD40L, a novel cytokine ligand for CD40.
 XX
 KW Cytokine; activated CD4+ T cell; CD40; CD40L; monoclonal antibody;
 KW neoplastic cell proliferation; B-cell lymphoma; immune deficiency; AIDS;
 KW melanoma; carcinoma.
 XX
 OS Homo sapiens.
 XX
 PN US5674492-A.
 XX
 PD 07-OCT-1997.

XX 21-DEC-1994: 94US-0360923.
 PF 23-DEC-1993: 93US-0172664.
 PR (IMMUNEX CORP.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Armitage RJ, Fanslow WC, Longo DL, Murphy WJ;
 XX WPI: 1997-502273/46.
 DR N-PSDB: AAT93782.
 PT Treating or preventing neoplastic disease associated with
 PT CD40-expressing cells - particularly B-cell lymphoma, by
 PT administration of CD40-binding protein, preferably antibody or
 PT soluble CD40-ligand
 XX
 PS Disclosure: Columns 21-24: 21pp; English.
 CC The present sequence represents the amino acid sequence of a novel
 CC cytokine ligand for CD40 called CD40L. CD40L is a type II membrane
 CC polypeptide, which is expressed by activated CD4+ T cells. It has an
 CC extracellular region at its C-terminus, and an intracellular region at
 CC which in turn bind to CD40-expressing cells. This inhibits binding of
 CC soluble CD40 to its ligand CD40L. The monoclonal antibody against CD40L
 CC is used to inhibit proliferation of neoplastic cells, and is particularly
 CC useful in treating B-cell lymphoma (e.g. where induced after transplants
 CC or in other cases of immune deficiency such as AIDS), and also melanoma
 CC or carcinoma. Since the monoclonal antibodies inhibit neoplastic cells
 CC directly, they may not need to be coupled to a toxin or radioisotope,
 CC avoiding toxic effects on normal B cells.
 CC
 XX Sequence 261 AA:
 SQ
 Query Match 44.6%; Score 1110; DB 18; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.3e-73;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 257 HRRLDKIEDERNHEDFVFMKTIORCNTGERSLSLNCERIKSOEGFVKDMLNKEETK 316
 DB 47 hrrldkiedernhedfvmktitqrntgerslsllnceelksqfegfvdmlnkeetk 106
 QY 317 KENSFEKQGDONPOIAAHVISEASRTTSVLOMAEKGYTMSNNLVTLNKGKOLTVKRO 376
 DB 107 kensfemqgdqnpqiaahviseasrttsvlgwaekgytmsnnlvtlengkqltvkrq 166
 QY 377 GLYYIYAOVTFCSNBRASSQAPFIASLCLKSPGREFRILLRAANTHSSAKPCGQOSIHIG 436
 DB 167 glyyyiyagvtfcsnreassqapfiaslclkspgferillraanthssakpcgqgsihlg 226
 QY 437 GFPELOPGASVFNVTDPDSOVSHGTFSGFLKL 471
 DB 227 gfpejpgasvfnvtdpsovshgthgftsfglkl 261
 RESULT 6
 AAW71751
 ID AAW71751 standard; Protein: 261 AA.
 AC AAW71751;
 XX
 XX 08-DEC-1998 (first entry)
 DE Human CD40 ligand.
 XX Human CD40 ligand.
 DE Human; CD40 ligand; TNF receptor family; activated T cell;
 KW type 2 membrane glycoprotein; cell proliferation; differentiation;
 KW B cell.
 XX Homo sapiens.
 OS

XX US5817516-A.
 PN 06-OCT-1998.
 PD 28-APR-1995: 95US-0431055.
 XX 28-APR-1995: 95US-0431055.
 PR 28-APR-1995: 95US-0431055.
 XX 28-APR-1994: 94US-0234580.
 PA (BOEH) BOEHRINGER INGELHEIM PHARM INC.
 XX Castle B, Kehry M;
 PI WPI: 1998-556393/47.
 DR N-PSDB: AAW61063.
 PT Increased proliferation of B cells in culture - by incubating them
 PT in the presence of membrane-bound CD40 ligand
 XX
 PS Example 1: Fig 1: 37pp; English.
 CC The present sequence represents human CD40 ligand which is used in the
 CC method of the invention. The method has been developed for proliferating
 CC B cells to increase their number at least 100-fold. The method
 CC comprises: (a) providing high density, membrane bound CD40 ligand; and
 CC (b) culturing one or more B cells in the presence of this ligand. The
 CC culture results in a proliferation in the number of B cells of at least
 CC 100 fold. Also described is a method as above where the B cells are
 CC induced to differentiate into antibody-producing cells in the presence
 CC of one or more cytokines. The method can be used for stimulating B-cell
 CC proliferation in vitro or in vivo, e.g. for treating conditions in which
 CC B-cell proliferation and activation is suppressed. Eight rounds of
 CC division over six days can be achieved, corresponding to a 256-fold
 CC increase in cell numbers, which is a vast increase compared to previous
 CC culturing methods.
 CC
 XX Sequence 261 AA:
 SQ
 Query Match 44.6%; Score 1110; DB 19; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.3e-73;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 257 HRRLDKIEDERNHEDFVFMKTIORCNTGERSLSLNCERIKSOEGFVKDMLNKEETK 316
 DB 47 hrrldkiedernhedfvmktitqrntgerslsllnceelksqfegfvdmlnkeetk 106
 QY 317 KENSFEKQGDONPOIAAHVISEASRTTSVLOMAEKGYTMSNNLVTLNKGKOLTVKRO 376
 DB 107 kensfemqgdqnpqiaahviseasrttsvlgwaekgytmsnnlvtlengkqltvkrq 166
 QY 377 GLYYIYAOVTFCSNBRASSQAPFIASLCLKSPGREFRILLRAANTHSSAKPCGQOSIHIG 436
 DB 167 glyyyiyagvtfcsnreassqapfiaslclkspgferillraanthssakpcgqgsihlg 226
 QY 437 GFPELOPGASVFNVTDPDSOVSHGTFSGFLKL 471
 DB 227 gfpejpgasvfnvtdpsovshgthgftsfglkl 261
 RESULT 7
 AAW41178
 ID AAW41178 standard; Protein: 261 AA.
 AC AAW41178;
 XX
 XX 13-MAY-1998 (first entry)
 DE CD40 ligand.
 XX Leucine zipper; fusion protein production; soluble oligomeric protein;
 KW heterologous mammalian type II transmembrane protein; activated T cell;
 KW

KM heterologous mammalian type I transmembrane protein; antibody production;
 KM CD40-L; B-cell proliferation; CD27-L; lymphocyte antigen.

XX Homo sapiens.
 XX US5716805-A.
 XX
 XX 10-FEB-1998.
 XX
 XX 18-MAY-1995; 95US-0446922.
 XX
 XX 18-MAY-1995; 95US-0446922.
 PR 18-MAY-1995; 95US-0446922.
 PR 25-OCT-1991; 91US-0783707.
 PR 05-DEC-1991; 91US-0805723.
 PR 23-OCT-1992; 92US-0969703.
 PR 13-AUG-1993; 93US-0107353.

XX (IMMV) IMMUNEX CORP.

PI Spriggs MK, Strinivasan S;

DR WPI; 1998-144799/13.

DR N-PSDB; AAV12852.

PT Soluble oligomeric fusion proteins - comprising leucine zipper fused
 to extracellular region of transmembrane protein

XX Example 1; column 21-22; 21pp; English.

XX This sequence represents the the human CD40 ligand (CD40-L).
 CC This protein can be used in a fusion protein produced using the
 CC method of the invention. The method is for preparing soluble oligomeric
 CC protein by culturing a host cell transfected with a vector for the
 CC fusion protein. The soluble oligomeric proteins comprise a leucine zipper
 CC fused to the N terminus of the extracellular region of a heterologous
 CC mammalian type II transmembrane protein or to the C terminus of the
 CC extracellular region of a heterologous mammalian type I transmembrane
 CC protein, where the leucine zipper is a peptide comprising at least part
 CC of AA41171 or AA41172, optionally with conservative amino acid
 CC substitutions, provided that the peptide trimerizes in solution. A
 CC soluble fusion protein comprising the leucine zipper of AA41171 linked
 CC to the extracellular region of CD40-L (a type II transmembrane protein
 CC that is found on activated T cells and acts as a ligand for the B-cell
 CC antigen CD40) stimulates B-cell proliferation and antibody production in
 CC a similar manner to membrane-bound CD40-L. A soluble fusion protein
 CC comprising the leucine zipper of AA41172 linked to the extracellular
 CC region of CD27-L (a type II transmembrane protein that binds to the
 CC lymphocyte antigen CD27) inhibits binding of CD27c (a fusion protein
 CC comprising the extracellular region of CD27 and a human IgG1 Fc region)
 CC to EBV-transformed B cells expressing CD27-L.

XX Sequence 261 AA;

Query Match 44.6%; Score 1110; DB 19; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.3e-73;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 HRRLDKIEDERNLHEDVFPMKTIORCNTGERSLSLNCCEIKSQFEGFVDMNKKEETK 316
 Db 47 hrrldkiedernlhedvfpmktiorcntgerslslnccceiksqfegfvdmnkkeetk 106
 OY 317 KENSFENQKGDQNPQIAAHVISEASSKTTSVLQWAEKGYTMSNNLVTLENGKQLTVKRQ 376
 Db 107 kensfemqkqdpnpqiaahviseasskttsvlqwaekgytmsnnlvtlengkqltvkrq 166
 OY 377 GLYIYIAQVTFCSNRKSSQAPFIASCLKSPGRFERILLRAANTHSSARPCGOOSTHIG 436
 Db 167 glyiyiaqvtfcsnrkssqapfiaslclksprferillraanthssarkpcgoostihig 226
 OY 437 GVFELOPGASVFNVTDPQSOVSHGCTGFTSGFLKL 471
 Db 227 gvfeilpgasvfnvtdpsqvsbgtgftsfglkl 261

RESULT 8

AAV39938
 ID AAV39938 standard; Protein; 261 AA.

XX AAV39938;

XX 13-DEC-1999 (first entry)

XX Human CD40-L protein sequence.

XX CD40-L; CD40 receptor ligand; anti-human CD40-L monoclonal antibody;
 KM binding inhibitor; trimeric CD40-L; anti-immunoglobulin M;
 KM peripheral blood B cell; proliferation inhibitor.

XX Homo sapiens.

XX US5961974-A.

XX 05-OCT-1999.

XX 24-MAY-1994; 94US-0249189.

XX 25-OCT-1991; 91US-0783707.

XX 05-DEC-1991; 91US-0805723.

XX 23-OCT-1992; 92US-0969703.

XX (IMMV) IMMUNEX CORP.

PI Spriggs MK, Fanslow WC, Armitage RJ;

DR WPI; 1999-579604/49.

DR N-PSDB; AA227525.

PT Anti-human CD40-Ligand monoclonal antibodies -

XX Disclosure; Fig 2; 59pp; English.
 PS This sequence represents the human CD40 receptor ligand (CD40-L). The
 CC invention relates to anti-human CD40-L monoclonal antibodies M90 secreted
 CC by hybridoma hCD40L-M90 (ATCC HB-12055) and M91 secreted by hybridoma
 CC hCD40L-M91 (ATCC HB-12056). M90 and M91 inhibit CD40-L binding to CD40
 CC and the ability of trimeric CD40-L and anti-immunoglobulin M to induce
 CC proliferation of peripheral blood B cells.

XX Sequence 261 AA;

Query Match 44.6%; Score 1110; DB 20; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.3e-73;
 Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 HRRLDKIEDERNLHEDVFPMKTIORCNTGERSLSLNCCEIKSQFEGFVDMNKKEETK 316
 Db 47 hrrldkiedernlhedvfpmktiorcntgerslslnccceiksqfegfvdmnkkeetk 106
 OY 317 KENSFENQKGDQNPQIAAHVISEASSKTTSVLQWAEKGYTMSNNLVTLENGKQLTVKRQ 376
 Db 107 kensfemqkqdpnpqiaahviseasskttsvlqwaekgytmsnnlvtlengkqltvkrq 166
 OY 377 GLYIYIAQVTFCSNRKSSQAPFIASCLKSPGRFERILLRAANTHSSARPCGOOSTHIG 436
 Db 167 glyiyiaqvtfcsnrkssqapfiaslclksprferillraanthssarkpcgoostihig 226
 OY 437 GVFELOPGASVFNVTDPQSOVSHGCTGFTSGFLKL 471
 Db 227 gvfeilpgasvfnvtdpsqvsbgtgftsfglkl 261
 RESULT 9
 AAV25155
 ID AAV25155 standard; protein; 261 AA.

AA25155;
27-AUG-1999 (first entry)
Human soluble CD40L protein.
CD40L; human; soluble; inhibitor; immune response; T-cell function; alloimmunization; immunotherapy; antibody response; model; PBL; Th cell; peripheral blood lymphocytes; therapy; B cell; APC; CD40; T-helper cell; cytokine secretion; transfusion-induced alloimmune response; anti-HLA; systemic lupus erythematosus; SLE; Sjogrens syndrome; Raynaud's syndrome; scleroderma myositis; type 1 diabetes; arthritis; rheumatoid arthritis; inflammatory bowel disease; uveitis; myasthenia gravis; allergy; multiple sclerosis; idiopathic thrombocytopenic purpura; antiprotective graft vs host disease; dermatologic; immunosuppressive; antidiabetic; antiinflammatory; antirheumatic; antiarthritic; antiallergic.
Homo sapiens.
MO9927948-A2.
10-JUN-1999.
27-NOV-1998; 98MO-CA01105.
28-NOV-1997; 97CA-2223225.
(CABL-) CANADIAN BLOOD SERVICES.]
CROW AR, Freedman J, Lazarus AR;
MPI; 1999-404800/34.
Use of recombinant human CD40L protein for treating immune diseases
Claim 2: Page 41-42; 42pp; English.
This invention describes a novel use of a soluble recombinant human CD40L (rhCD40L) protein or protein fragment (containing the active binding site with CD40) for inhibiting an immune response and T-cell function. The invention also describes (1) an immunodeficient mouse model of human alloimmunization for testing in vivo effects of an immunotherapy or inhibition of a human antibody response, where the mouse model is an immunodeficient mouse reconstituted with human peripheral blood lymphocytes (PBL) from donors and (2) A method for inhibiting an immune response, or T-cell function, in a patient by administering a therapeutic amount of rhCD40L or functional, in a patient by administering a therapeutic amount of rhCD40L or functional fragment. The soluble recombinant CD40L interaction with CD40L on the Th (T-helper) cells which disallows the Th cell to be activated to secrete cytokines (such as Th2 cytokines) which thus reduce the transfusion-induced alloimmune response. The rhCD40L or its fragment is useful for inhibiting a human anti-HLA alloimmune response. Where T-cell function is inhibited the following diseases can be treated or prevented: systemic lupus erythematosus (SLE), Sjogrens syndrome, scleroderma myositis, Raynaud's syndrome, type 1 diabetes, arthritis and rheumatoid arthritis, inflammatory bowel disease, uveitis, myasthenia gravis, multiple sclerosis, idiopathic thrombocytopenic purpura and graft vs host disease, and allergies dependent on T-cells. The development of the model will allow detailed study of the mechanisms of alloimmunization and allow the development of new strategies for the modulation of human alloimmunization to blood cell antigens. Administration of a recombinant 18kDa-CD154 molecule can inhibit an alloimmune response. This molecule may have good therapeutic potential to inhibit human transfusion-induced alloimmunization. The product of antirheumatic, antiarthritic, immunosuppressive, antiinflammatory, antiallergic activity.

| | Best Local Similarity | 100.0% | Pred. No. 1.3e-73: | |
|---------|-----------------------|---|--------------------|------------|
| Matches | 215: | Conservative | 0: | Mismatches |
| | | | 0: | Indels |
| | | | | 0: |
| | | | | Gaps |
| | | | | 0: |
| QY | 257 | HRRLDKIEDERNLHEEFVFKTITQRCNCTGSLSLNCEETKSQEGFVKYDMLNKEETK | 316 | |
| | | | | |
| Db | 47 | hrrldkiedernlhnelfvfmktlqrcnctgerslslnceelksqefgfykdlmlnkeetk | 106 | |
| QY | 317 | KENSEFMOKGDONPOTIAAHVISEASKTTSYLQMAEKGYTSMNLVLTENGKOLTVRQ | 376 | |
| | | | | |
| Db | 107 | kensfemgqdgdpqjaalvaivisaasktsvlyvqekgytmsnmlvltengkqltvrrq | 166 | |
| QY | 377 | GLYLYYAOVTEFCNSNEBASSQAPFIASLCKSGGREERILLRAANHSASAKCCGOOSHLG | 436 | |
| | | | | |
| Db | 167 | glylyyaqvtfcsnreassgapflaslckpgrferrlllraanhsasakpcgqqslnhg | 226 | |
| QY | 437 | GVEELQPGASVYVNTDPSOYSHGCGFISFGILKL | 471 | |
| | | | | |
| Db | 227 | gvfelqpgasvrvnvtldpsqyvhngcgflsfgilkl | 261 | |

RESULT 10
ID AAY96993 standard; Protein; 261 AA.
XN AAY96993;
XN AAY96993;
XN 31-OCT-2000 (first entry)
XN
XN Human CD40 ligand.
XN
XN CD40 ligand; CD40; T cell; T cell receptor; rearrangement; maturation;
XN cell death inhibition; stress-induced; immunosuppressive; anti-thyroid;
XN anti-inflammatory; anti-diabetic; anti-rheumatic; anti-anaemic;
XN ophthalmological; anti-psoriatic; nephrotoxic; hepatotropic; virucide;
XN dermatological; cytostatic.
XN
XN Homo sapiens.
XN
XN WO200039283-A1.
XN
XN 06-JUL-2000.
XN
XN 22-DEC-1999; 99WO-US30930.
XN
XN 29-DEC-1998; 98US-0114106.
XN
XN (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
XN
XN Newell MK, Wagner D, Newell E;
XN
XN WPI: 2000-452387/39.
XN
XN N-PSDB: AAA51745.
XN
XN Inducing T cell receptor gene rearrangement for treating autoimmune
XN diseases comprises contacting T cells with a CD40-binding agent
XN
XN Disclosure; Page 47; 50pp; English.
XN
XN
XN CD40 engagement on T cells can be used to induce T cell receptor
XN rearrangement and enhance T cell affinity for a particular antigen. CD40
XN engagement can be bought about by contacting CD40 with a CD40-binding
XN agent, e.g. human CD40 ligand. The CD40-binding agents can also be used
XN in methods for promoting T cell maturation, inhibiting T cell receptor
XN rearrangement, inhibiting environmental stress-induced cell death,
XN altering the specificity of a T cell towards an antigen, inducing T cell
XN reactivity towards an antigen or enhancing environmental stress-induced
XN cell death (all claimed). T cell affinity maturation towards a specific
XN antigen can be inhibited, especially for a self-antigen in an autoimmune
XN disease, which includes rheumatoid arthritis, uveitis, insulin-dependent
XN diabetes mellitus, haemolytic anaemias, rheumatic fever, Crohn's disease
XN Guillain-Barre syndrome, psoriasis, thyroiditis, Grave's disease,
XN myasthenia gravis, glomerulonephritis, autoimmune hepatitis or systemic
XN lupus erythematosus. Inducing environmental stress-induced T cell death

CC is carried out in a cancerous T cell or a self-reactive T cell where the
CC environmental stress is a chemotherapeutic agent (claimed).
XX
SQ Sequence 261 AA:

Query Match 44.6%; Score 1110; DB 21; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.3e-73;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 HRRLDKIEDERNHEDVFVKTTORCNTGERSLSLNCCEIKSQFEGFVADIMLNKEETK 316
DB 47 hrrldkiedernhedefvfkttqrcntgterslsllnceelkqfegfvadimlnkeetk 106
OY 317 KENSFEKQKDONPOIAAHVISEASSKTTSVLOWAEKGYTMSNNLTLENGKQLTVKRO 376
DB 107 kensfemqkqdonpqiaahviseasskttavlwaeqgytmsnnltlengkqltvkrq 166
OY 377 GLYYIYAQVTFCSNREASSQAPFIASCLKSPGRFERILLRAANTHSSAKPCGQOSIHLG 436
DB 167 glyyyiaqvtfcsnreassqapfiassclksprferillraanthssakpcgqqsihlg 226
OY 437 GFVELQPGASVFNWVTPSOVSHGTGFTSPGLKL 471
DB 227 gfvelqpgasvfnwvtpsovshgtgftspglkl 261

RESULT 11
AAB67612
ID AAB67612 standard; protein; 261 AA.

XX AAB67612:

DT 29-MAY-2001 (first entry)

DE Amino acid sequence of human gp39 protein, a CD40 ligand.

XX gp39; CD40 ligand; osteoblast cell death; apoptosis; bone loss;
KM osteoporosis; osteonecrosis; inflammatory arthritis; estrogen loss;
KW ovariectomy; hysterectomy; lupus nephritis; Takayasu's arteritis;
KM Wegener's granulomatosis; nephritis; myositis; scleroderma;
KW thrombocytopenia; asthma; lung disease; cancer.

OS Homo sapiens.

XX WO200116180-A2.

PD 08-MAR-2001.

PF 24-AUG-2000; 2000MO-US23276.

PR 27-AUG-1999; 99US-0151250.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Ahuja SS, Bonewald LF;

DR WPI; 2001-169007/17.

DR N-PSDB; AAF55539.

PT CD40 agonist containing composition, used to reduce bone cell death or
PT apoptosis associated with osteoporosis, osteonecrosis and inflammatory
PT arthritis -

PS Claim 3; Page 113-114; 118pp; English.

XX The present sequence represents a gp39 protein. It is a CD40 ligand.
CC CD40 ligands are used for reducing osteoblast cell death or apoptosis,
CC and for treating or preventing bone loss in animals, preferably humans,
CC at risk of, or undergoing, bone loss. The bone loss is associated with
CC osteoporosis, osteonecrosis, inflammatory arthritis, post-menopausal
CC estrogen loss, estrogen loss due to ovariectomy, total hysterectomy,
CC lupus nephritis, Takayasu's arteritis, Wegener's granulomatosis,

CC anti-glomerular basement membrane nephritis, myositis, scleroderma,
CC idiopathic autoimmune thrombocytopenia, asthma, a chronic obstructive
CC lung disease, nephrotic/nephritic syndrome, or cancer. They may also be
CC used to treat or prevent bone loss in a subject undergoing, or scheduled
CC for, an organ or bone marrow transplant.

XX
SQ Sequence 261 AA:

Query Match 44.6%; Score 1110; DB 22; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.3e-73;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 HRRLDKIEDERNHEDVFVKTTORCNTGERSLSLNCCEIKSQFEGFVADIMLNKEETK 316
DB 47 hrrldkiedernhedefvfkttqrcntgterslsllnceelkqfegfvadimlnkeetk 106
OY 317 KENSFEKQKDONPOIAAHVISEASSKTTSVLOWAEKGYTMSNNLTLENGKQLTVKRO 376
DB 107 kensfemqkqdonpqiaahviseasskttavlwaeqgytmsnnltlengkqltvkrq 166
OY 377 GLYYIYAQVTFCSNREASSQAPFIASCLKSPGRFERILLRAANTHSSAKPCGQOSIHLG 436
DB 167 glyyyiaqvtfcsnreassqapfiassclksprferillraanthssakpcgqqsihlg 226
OY 437 GFVELQPGASVFNWVTPSOVSHGTGFTSPGLKL 471
DB 227 gfvelqpgasvfnwvtpsovshgtgftspglkl 261

RESULT 12

AAB37806
ID AAB37806 standard; protein; 261 AA.

XX AAB37806:

DT 23-FEB-2001 (first entry)

DE Human CD40 ligand sequence.

XX Human CD40 ligand; antiviral; anti-HIV; CD40 stimulant;

KM human immunodeficiency virus; HIV; cancer; organ transplantation.

XX Homo sapiens.

XX WO200066155-A1.

PD 09-NOV-2000.

PF 28-APR-2000; 2000MO-US11734.

PR 30-APR-1999; 99US-0131730.

PA (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.

PI Sarawar SR, Schoenberger SP;

DR WPI; 2001-024740/03.

PT Method for preventing latent virus reactivation or controlling virus
PT replication by administration of antibodies or similar reagents which
PT stimulate the CD40 molecule -

PS Disclosure; Page 16; 36pp; English.

XX The present sequence is a human CD40 ligand. It may be used in a method
CC for preventing latent virus reactivation or controlling virus
CC replication. The method comprises administering a composition capable
CC of binding to a cell membrane expressed CD40, where the binding of the
CC composition to the CD40 on the surface of the cell generates a
CC stimulatory signal to the cell. The method is useful for treating
CC immunocompromised patients, e.g. infected with human immunodeficiency
CC virus (HIV), or immunosuppressed patients, e.g. due to infections or

CC illness, or as a side effect of treatments (such as radiation or cancer
 CC therapy), or due to tissue or organ transplantation. The method does
 CC not have the levels of toxicity seen following administration of
 CC conventional antiviral drugs.
 CC
 XX Sequence 261 AA:

Query Match 44.6%: Score 1110: DB 22: Length 261;
 Best Local Similarity 100.0%: Pred. No. 1.3e-73;
 Matches 215: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

QY 257 HRLDKIEDERLHEDFVFMKTQRCNTGERSLNLNCEERKSQEGFVKDMLNKEETK 316
 DB 47 HRLDKIEDERLHEDFVFMKTQRCNTGERSLNLNCEERKSQEGFVKDMLNKEETK 106
 QY 317 KENFEMOKGQDNPQIAAHVISEASSKTSVLOMAEKGYTMSNNLVLENGKQLTVKRQ 376
 DB 107 KENFEMOKGQDNPQIAAHVISEASSKTSVLOMAEKGYTMSNNLVLENGKQLTVKRQ 166
 QY 377 GLYTYIAQVTFCSNREASSQAPFIASLCLKSPGFRERILLRAANTHSSAKPCGOOSIHLG 436
 DB 167 GLYTYIAQVTFCSNREASSQAPFIASLCLKSPGFRERILLRAANTHSSAKPCGOOSIHLG 226
 QY 437 GVFELOPGASVFNVTDPGSHGTFSGFLKL 471
 DB 227 GVFELOPGASVFNVTDPGSHGTFSGFLKL 261

RESULT 13

AAW09122
 ID AAW09122 standard; Protein: 294 AA.

XX AAW09122:

XX 24-SEP-1997 (first entry)

DE CD40 ligand/zipper domain fusion protein mutant 41.

KW CD40 ligand: membrane bound glycoprotein; B cell proliferation;
 antibody secretion; immunoglobulin E; cytokine; CD40L; mutin;

KM oligomerisation domain; fusion protein; dimer; trimer.

XX Chimeric - Homo sapiens.

OS Chimeric - Saccharomyces cerevisiae.

XX Synthetic.

XX Key

FT Domain

FT Location/Qualifiers

FT 1..33

FT /label= leucine zipper_15M_116T

FT /note= "Mutant version of yeast GCN4 leucine zipper"

FT Misc-difference 5

FT /note= "Wild-type Ile has been replaced by Met"

FT /note= "Wild-type Ile has been replaced by Thr"

FT Region

FT 34..294

FT /label= CD40L

XX WO9640918-A2.

XX 19-DEC-1996.

XX New CD40 ligand mutin with higher CD40 affinity than native ligand
 PT - useful in binding assays, and for therapy of disorders and
 PT diseases involving low levels of B cells and antibody secretion
 PS
 XX Example 4: Page -; 31pp: English.
 CC This sequence represents a mutant version of human CD40 ligand/
 CC zipper domain fusion protein which is able to bind CD40 in a solid
 CC phase binding assay. The random mutations were generated by PCR
 CC misincorporation into yeast expression constructs and mutants were
 CC selected based on an apparent increase in secretion.
 CC (Note: The present sequence does not appear in the specification;
 CC it has been produced using the sequences of the leucine zipper and
 CC wild-type CD40L which are given on page 23 and pages 20-21,
 CC respectively).
 XX
 SQ Sequence 294 AA:

Query Match 44.6%: Score 1110: DB 18: Length 294;
 Best Local Similarity 100.0%: Pred. No. 1.5e-73;
 Matches 215: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

QY 257 HRLDKIEDERLHEDFVFMKTQRCNTGERSLNLNCEERKSQEGFVKDMLNKEETK 316
 DB 80 HRLDKIEDERLHEDFVFMKTQRCNTGERSLNLNCEERKSQEGFVKDMLNKEETK 139
 QY 317 KENFEMOKGQDNPQIAAHVISEASSKTSVLOMAEKGYTMSNNLVLENGKQLTVKRQ 376
 DB 140 KENFEMOKGQDNPQIAAHVISEASSKTSVLOMAEKGYTMSNNLVLENGKQLTVKRQ 199
 QY 377 GLYTYIAQVTFCSNREASSQAPFIASLCLKSPGFRERILLRAANTHSSAKPCGOOSIHLG 436
 DB 200 GLYTYIAQVTFCSNREASSQAPFIASLCLKSPGFRERILLRAANTHSSAKPCGOOSIHLG 259
 QY 437 GVFELOPGASVFNVTDPGSHGTFSGFLKL 471
 DB 260 GVFELOPGASVFNVTDPGSHGTFSGFLKL 294

RESULT 14

AAW09119
 ID AAW09119 standard; Protein: 294 AA.

XX AAW09119:

XX 24-SEP-1997 (first entry)

DE CD40 ligand/zipper domain fusion protein mutant 14.

KW CD40 ligand: membrane bound glycoprotein; B cell proliferation;
 antibody secretion; immunoglobulin E; cytokine; CD40L; mutin;

KM oligomerisation domain; fusion protein; dimer; trimer.

XX Chimeric - Homo sapiens.

OS Chimeric - Saccharomyces cerevisiae.

XX Synthetic.

XX Key

FT Domain

FT Location/Qualifiers

FT 1..33

FT /label= leucine zipper_112N

FT /note= "Mutant version of yeast GCN4 leucine zipper"

FT Misc-difference 12

FT /note= "Wild-type Ile has been replaced by Asn"

FT Region

FT 34..294

FT /label= CD40L_K260N

FT /note= "Mutant version of CD40 ligand"

FT Misc-difference 293

FT /note= "Wild-type Lys has been replaced by Asn"

FT WO9640918-A2.

Tue Jul 3 16:26:03 2001

us-09-454-223-2.rag

Page 10

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:13:43 ; Search time 17.98 seconds

(without alignments)
527.704 Million cell updates/sec

Title: US-09-454-223-2

Perfect score: 2487
Sequence: 1 MLPLSMVLVLVPLGNLGA.....TDPQVSHGTSTGSLKL 471

Scoring table:

BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 193259 seqs, 20144639 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/prodata/2/1aa/5a_COMB.pep: *
2: /cgn2_6/prodata/2/1aa/5b_COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6a_COMB.pep: *
4: /cgn2_6/prodata/2/1aa/5b_COMB.pep: *
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/prodata/2/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------------|
| 1 | 1110 | 44.6 | 261 | 1 | US-07-940-605A-2 |
| 2 | 1110 | 44.6 | 261 | 1 | US-08-184-422-8 |
| 3 | 1110 | 44.6 | 261 | 1 | US-08-360-923A-2 |
| 4 | 1110 | 44.6 | 261 | 1 | US-08-446-922-4 |
| 5 | 1110 | 44.6 | 261 | 2 | US-08-431-055-4 |
| 6 | 1110 | 44.6 | 261 | 2 | US-08-690-096-2 |
| 7 | 1110 | 44.6 | 261 | 2 | US-08-249-189-12 |
| 8 | 1110 | 44.6 | 261 | 2 | US-08-484-624A-12 |
| 9 | 1110 | 44.6 | 261 | 2 | US-08-477-733B-12 |
| 10 | 1110 | 44.6 | 261 | 3 | US-08-763-995-2 |
| 11 | 1110 | 44.6 | 261 | 3 | US-09-088-913A-12 |
| 12 | 1110 | 44.6 | 261 | 3 | US-08-589-771B-8 |
| 13 | 1110 | 44.6 | 261 | 5 | PCT-US93-10034-4 |
| 14 | 1094.5 | 44.0 | 273 | 1 | US-08-446-922-11 |
| 15 | 1094.5 | 44.0 | 273 | 2 | US-08-249-189-21 |
| 16 | 1094.5 | 44.0 | 273 | 2 | US-08-484-624A-21 |
| 17 | 1094.5 | 44.0 | 273 | 2 | US-08-477-733B-21 |
| 18 | 1094.5 | 44.0 | 273 | 3 | US-09-088-913A-21 |
| 19 | 1089 | 43.8 | 473 | 2 | US-08-249-189-16 |
| 20 | 1089 | 43.8 | 473 | 2 | US-08-484-624A-16 |
| 21 | 1089 | 43.8 | 473 | 2 | US-08-477-733B-16 |
| 22 | 1089 | 43.8 | 473 | 3 | US-09-088-913A-16 |
| 23 | 831 | 33.4 | 260 | 1 | US-08-446-922-6 |
| 24 | 831 | 33.4 | 260 | 2 | US-08-249-189-2 |
| 25 | 831 | 33.4 | 260 | 2 | US-08-484-624A-2 |
| 26 | 831 | 33.4 | 260 | 2 | US-08-477-733B-2 |
| 27 | 831 | 33.4 | 260 | 3 | US-09-088-913A-2 |

| | | | | | | |
|----|-------|------|------|---|-------------------|-------------------|
| 28 | 831 | 33.4 | 260 | 5 | PCT-US93-10034-6 | Sequence 6, App1 |
| 29 | 829.5 | 33.3 | 351 | 4 | US-09-011-735-1 | Sequence 1, App1 |
| 30 | 827 | 33.3 | 260 | 2 | US-08-431-055-2 | Sequence 2, App1 |
| 31 | 812 | 32.6 | 280 | 2 | US-08-249-189-23 | Sequence 23, App1 |
| 32 | 812 | 32.6 | 280 | 2 | US-08-484-624A-23 | Sequence 23, App1 |
| 33 | 812 | 32.6 | 280 | 2 | US-08-477-733B-23 | Sequence 23, App1 |
| 34 | 812 | 32.6 | 280 | 3 | US-09-088-913A-23 | Sequence 23, App1 |
| 35 | 766 | 30.8 | 149 | 3 | US-08-584-031-16 | Sequence 16, App1 |
| 36 | 755 | 30.4 | 151 | 1 | US-07-940-605A-3 | Sequence 3, App1 |
| 37 | 755 | 30.4 | 151 | 1 | US-08-690-096-3 | Sequence 3, App1 |
| 38 | 652 | 26.2 | 171 | 4 | US-09-011-735-2 | Sequence 2, App1 |
| 39 | 567 | 22.8 | 151 | 2 | US-07-940-605A-4 | Sequence 4, App1 |
| 40 | 567 | 22.8 | 151 | 2 | US-08-690-096-4 | Sequence 4, App1 |
| 41 | 481.5 | 19.4 | 1057 | 3 | US-08-931-820-1 | Sequence 1, App1 |
| 42 | 474.5 | 19.1 | 1057 | 3 | US-08-931-820-4 | Sequence 4, App1 |
| 43 | 474.5 | 19.1 | 1341 | 4 | US-08-963-825-18 | Sequence 18, App1 |
| 44 | 473.5 | 19.0 | 1060 | 3 | US-08-931-820-3 | Sequence 3, App1 |
| 45 | 473.5 | 19.0 | 1418 | 4 | US-08-963-825-20 | Sequence 20, App1 |

ALIGNMENTS

RESULT 1
US-07-940-605A-2
Sequence 2, Application US/07940605A
Patent No. 5540926
GENERAL INFORMATION:
APPLICANT: ARUFPO, ALEJANDRO
APPLICANT: HOLLENBAUGH, DIANE
APPLICANT: LEDBETTER, JEFFREY A.
TITLE OF INVENTION: SOLUBLE LIGANDS FOR CD40
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/940, 605A
FILING DATE: 04-SEP-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-184
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-940-605A-2

Query Match 44.68; Score 1110; DB 1; Length 261;
Best Local Similarity 100.04; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 HRRLDKIEDERNLHEDFVFMKTIORCNTGERSLSLNCCEIKSOFEGFVMDIMLNKEETK 316
|||||
Db 47 HRRLDKIEDERNLHEDFVFMKTIORCNTGERSLSLNCCEIKSOFEGFVMDIMLNKEETK 106

0Y 317 KENSFEMQKGDONPOIAAHVISEASSTTSVLOMAEKGYTMSNNLVLENGKOLTVKRQ 376
DB 107 KENSFEMQKGDONPOIAAHVISEASSTTSVLOMAEKGYTMSNNLVLENGKOLTVKRQ 166
0Y 377 GLYTYIAQVTFCSNREASSOAPFIASLCLKSPGRFERILLRAANTHSSAKPCGOOSIHG 436
DB 167 GLYTYIAQVTFCSNREASSOAPFIASLCLKSPGRFERILLRAANTHSSAKPCGOOSIHG 226
0Y 437 GVFELOPGASVFVNWDPDSOVSHGTGTFSGLLKL 471
DB 227 GVFELOPGASVFVNWDPDSOVSHGTGTFSGLLKL 261

RESULT 2

US-08-184-422-8
Sequence 8, Application US/08184422
Patent No. 5565321
GENERAL INFORMATION:
APPLICANT: ARMITAGE, RICHARD
APPLICANT: DAVISON, BARRY
APPLICANT: FANSLAW, WILLIAM
APPLICANT: KENSHAW, WILLIAM
APPLICANT: SPRIGGS, MELANIE
APPLICANT: WIDMER, MICHAEL
TITLE OF INVENTION: DETECTION AND TREATMENT OF MUTATIONS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: Apple Macintosh
SOFTWARE: MS Word for Apple 5.1, Version a
CURRENT APPLICATION DATA:
FILING DATE: US/08/184, 422
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/009,258
FILING DATE: 01/22/93
ATTORNEY/AGENT INFORMATION:
NAME: PERKINS, PATRICIA ANNE
REGISTRATION NUMBER: 34,693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2065870430
TELEFAX: 2065870606
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-184-422-8

Query Match 44.6%; Score 1110; DB 1; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0Y 257 HRLDKIEDERNLHDEFFMKTICRCNTGERSLSLNCETIKSQEGFVXKIMLNKEETK 316
DB 47 HRLDKIEDERNLHDEFFMKTICRCNTGERSLSLNCETIKSQEGFVXKIMLNKEETK 106
0Y 317 KENSFEMQKGDONPOIAAHVISEASSTTSVLOMAEKGYTMSNNLVLENGKOLTVKRQ 376
DB 107 KENSFEMQKGDONPOIAAHVISEASSTTSVLOMAEKGYTMSNNLVLENGKOLTVKRQ 166

DB 107 KENSFEMQKGDONPOIAAHVISEASSTTSVLOMAEKGYTMSNNLVLENGKOLTVKRQ 166
0Y 377 GLYTYIAQVTFCSNREASSOAPFIASLCLKSPGRFERILLRAANTHSSAKPCGOOSIHG 436
DB 167 GLYTYIAQVTFCSNREASSOAPFIASLCLKSPGRFERILLRAANTHSSAKPCGOOSIHG 226
0Y 437 GVFELOPGASVFVNWDPDSOVSHGTGTFSGLLKL 471
DB 227 GVFELOPGASVFVNWDPDSOVSHGTGTFSGLLKL 261

RESULT 3

US-08-360-923A-2
Sequence 2, Application US/08360923A
Patent No. 5674492
GENERAL INFORMATION:
APPLICANT: ARMITAGE, RICHARD
APPLICANT: FANSLAW, WILLIAM
APPLICANT: LONGO, DAN L.
APPLICANT: MORPHY, WILLIAM
TITLE OF INVENTION: METHOD OF PREVENTING OR TREATING
TITLE OF INVENTION: DISEASE CHARACTERIZED BY NEOPLASTIC CELLS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: Apple Macintosh
SOFTWARE: Microsoft Word for Macintosh, Version 4.5.1a
CURRENT APPLICATION DATA:
FILING DATE: December 21, 1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/172,664
FILING DATE: December 23, 1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)587-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-360-923A-2

Query Match 44.6%; Score 1110; DB 1; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0Y 257 HRLDKIEDERNLHDEFFMKTICRCNTGERSLSLNCETIKSQEGFVXKIMLNKEETK 316
DB 47 HRLDKIEDERNLHDEFFMKTICRCNTGERSLSLNCETIKSQEGFVXKIMLNKEETK 106
0Y 317 KENSFEMQKGDONPOIAAHVISEASSTTSVLOMAEKGYTMSNNLVLENGKOLTVKRQ 376
DB 107 KENSFEMQKGDONPOIAAHVISEASSTTSVLOMAEKGYTMSNNLVLENGKOLTVKRQ 166
0Y 377 GLYTYIAQVTFCSNREASSOAPFIASLCLKSPGRFERILLRAANTHSSAKPCGOOSIHG 436
DB 107 KENSFEMQKGDONPOIAAHVISEASSTTSVLOMAEKGYTMSNNLVLENGKOLTVKRQ 166

Db 167 GLYIYAQVTFCSNREASAPFIASCLKSPGFERILLRAANTHSSARPCGOOSIHLG 226
QY 437 GFELQPGASVFNVTDPDSVSHGTGTFSGLLKL 471
Db 227 GFELQPGASVFNVTDPDSVSHGTGTFSGLLKL 261

RESULT 4

US-08-446-922-4
Sequence 4, Application US/08446922
Patent No. 5716805
GENERAL INFORMATION:
APPLICANT: Spriggs, Melanie
TITLE OF INVENTION: Methods of Preparing Soluble, Oligomeric
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Operating System 7.1
SOFTWARE: Microsoft Word for Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,922
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/107,353
FILING DATE: 08-13-93
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 1003-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-922-4

Query Match 44.6%; Score 1110; DB 1; Length 261;
Best local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 HRLDKIEDERNLHEDVFVKTIORCNTGERSLSLNCCEIKSQFEGFVDMLNKEETK 316
Db 47 HRLDKIEDERNLHEDVFVKTIORCNTGERSLSLNCCEIKSQFEGFVDMLNKEETK 106
QY 317 KENSFEMQKDONPOLAHYISEASSTTSVLOMAEKGYTMSNNLYTLENGKOLTVKRO 376
Db 107 KENSFEMQKDONPOLAHYISEASSTTSVLOMAEKGYTMSNNLYTLENGKOLTVKRO 166
QY 377 GLYIYAQVTFCSNREASAPFIASCLKSPGFERILLRAANTHSSARPCGOOSIHLG 436
Db 167 GLYIYAQVTFCSNREASAPFIASCLKSPGFERILLRAANTHSSARPCGOOSIHLG 226
QY 437 GFELQPGASVFNVTDPDSVSHGTGTFSGLLKL 471
Db 227 GFELQPGASVFNVTDPDSVSHGTGTFSGLLKL 261

RESULT 5

US-08-431-055-4
Sequence 4, Application US/08431055
Patent No. 5817516
GENERAL INFORMATION:
APPLICANT: KEHR, MERILYN R
TITLE OF INVENTION: METHODS FOR PROLIFERATING AND
DIFFERENTIATING B CELLS, AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX
STREET: 100 NEW YORK AVE. N.W. SUITE 600
CITY: WASHINGTON
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,055
FILING DATE: 28-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/234,580
FILING DATE: 28-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 1011.1030000/RAM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-431-055-4

Query Match 44.6%; Score 1110; DB 2; Length 261;
Best local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 HRLDKIEDERNLHEDVFVKTIORCNTGERSLSLNCCEIKSQFEGFVDMLNKEETK 316
Db 47 HRLDKIEDERNLHEDVFVKTIORCNTGERSLSLNCCEIKSQFEGFVDMLNKEETK 106
QY 317 KENSFEMQKDONPOLAHYISEASSTTSVLOMAEKGYTMSNNLYTLENGKOLTVKRO 376
Db 107 KENSFEMQKDONPOLAHYISEASSTTSVLOMAEKGYTMSNNLYTLENGKOLTVKRO 166
QY 377 GLYIYAQVTFCSNREASAPFIASCLKSPGFERILLRAANTHSSARPCGOOSIHLG 436
Db 167 GLYIYAQVTFCSNREASAPFIASCLKSPGFERILLRAANTHSSARPCGOOSIHLG 226
QY 437 GFELQPGASVFNVTDPDSVSHGTGTFSGLLKL 471
Db 227 GFELQPGASVFNVTDPDSVSHGTGTFSGLLKL 261

RESULT 6

US-08-690-096-2
Sequence 2, Application US/08690096
Patent No. 5945513
GENERAL INFORMATION:
APPLICANT: ARUFFO, ALEJANDRO

APPLICANT: HOLLENBAUGH, DIANE
APPLICANT: LEDBETTER, JEFFREY A.
TITLE OF INVENTION: SOLUBLE LIGANDS FOR CD40
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/690,096
FILING DATE: 31-JUL-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/940,605
FILING DATE: 04-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-184
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELLEX: 66141 PENNTE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-690-096-2

Query Match 44.6%; Score 1110; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 HRLDKIEDERNLHEDFVFMKTIORCNTGERSLSLNCCEIKSQFEGFYKDMLNKEETK 316
DB 47 HRLDKIEDERNLHEDFVFMKTIORCNTGERSLSLNCCEIKSQFEGFYKDMLNKEETK 106
QY 317 KENSFEMOKGQDNPQIAAHVISEASSKTTSVLQMAEKGYTMSNNLVTLNKGKQLTVKRO 376
DB 107 KENSFEMOKGQDNPQIAAHVISEASSKTTSVLQMAEKGYTMSNNLVTLNKGKQLTVKRO 166
QY 377 GLYYIAQVTFCSNREASSQAPFIASLCLKSPGRFERILLRAANTHSSAKPCGQOSIHIG 436
DB 167 GLYYIAQVTFCSNREASSQAPFIASLCLKSPGRFERILLRAANTHSSAKPCGQOSIHIG 226
QY 437 GVFELOPGASVFNVTDPDSQVSHGTGFTSGFLKL 471
DB 227 GVFELOPGASVFNVTDPDSQVSHGTGFTSGFLKL 261

RESULT 7
US-08-249-189-12
Sequence 12, Application US/08249189
Patent No. 5961974
GENERAL INFORMATION:
APPLICANT: ARMITAGE, RICHARD
APPLICANT: FANSLAW, WILLIAM
APPLICANT: SPRIGGS, MELANIE
APPLICANT: SRINIVASAN, SUBHASHINI
APPLICANT: GIBSON, MARYLOU
TITLE OF INVENTION: NOVEL CYTOKINE
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Operating System 7.1
SOFTWARE: Microsoft Word for Apple, version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/249,189
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/969,703
FILING DATE: October 23, 1992
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/783,707
FILING DATE: October 25, 1991
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2802-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2065870430
TELEFAX: 2065870606
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-249-189-12

Query Match 44.6%; Score 1110; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 HRLDKIEDERNLHEDFVFMKTIORCNTGERSLSLNCCEIKSQFEGFYKDMLNKEETK 316
DB 47 HRLDKIEDERNLHEDFVFMKTIORCNTGERSLSLNCCEIKSQFEGFYKDMLNKEETK 106
QY 317 KENSFEMOKGQDNPQIAAHVISEASSKTTSVLQMAEKGYTMSNNLVTLNKGKQLTVKRO 376
DB 107 KENSFEMOKGQDNPQIAAHVISEASSKTTSVLQMAEKGYTMSNNLVTLNKGKQLTVKRO 166
QY 377 GLYYIAQVTFCSNREASSQAPFIASLCLKSPGRFERILLRAANTHSSAKPCGQOSIHIG 436
DB 167 GLYYIAQVTFCSNREASSQAPFIASLCLKSPGRFERILLRAANTHSSAKPCGQOSIHIG 226
QY 437 GVFELOPGASVFNVTDPDSQVSHGTGFTSGFLKL 471
DB 227 GVFELOPGASVFNVTDPDSQVSHGTGFTSGFLKL 261

RESULT 8
US-08-484-624A-12
Sequence 12, Application US/08484624A
Patent No. 5962406
GENERAL INFORMATION:
APPLICANT: ARMITAGE, RICHARD
APPLICANT: FANSLAW, WILLIAM
APPLICANT: SPRIGGS, MELANIE

APPLICANT: SRINIVASAN, SUBHASHINI
APPLICANT: GIBSON, MARYLOU
APPLICANT: MORRIS, ARVIA E.
APPLICANT: MCGREW, JEFFERY
TITLE OF INVENTION: NOVEL CYTOKINE THAT BINDS CD40
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: MS Word for Apple Power Macintosh, version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,624A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/477,733
FILING DATE: June 07, 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/969,703
FILING DATE: October 23, 1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/805,723
FILING DATE: December 5, 1991
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/783,707
FILING DATE: October 25, 1991
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2802-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2065870430
TELEFAX: 2065870606
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-624A-12

Query Match 44.6%; Score 1110; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

257 HRRLDKIEDERNLHEDVFVKTIORCNTGERSLSLNCCEIKSOFEGFVNDIMLNKEETK 316
47 HRRLDKIEDERNLHEDVFVKTIORCNTGERSLSLNCCEIKSOFEGFVNDIMLNKEETK 106
317 KENSFENQKDONQIAAHYISEASSKTSVLQMAEKGYTMSNNLVTLENGQLTVKRQ 376
107 KENSFENQKDONQIAAHYISEASSKTSVLQMAEKGYTMSNNLVTLENGQLTVKRQ 166
377 GLYYIAQVYFCGNSRREASSAPFIASLQKSPGFERILLRANTHSSAKPCGOOSTHIG 436
167 GLYYIAQVYFCGNSRREASSAPFIASLQKSPGFERILLRANTHSSAKPCGOOSTHIG 226
437 GVEELDPGASVFNVTDPDSQVSHGTGFTSFGLLKL 471
227 GVEELDPGASVFNVTDPDSQVSHGTGFTSFGLLKL 261

RESULT 9
US-08-477-733B-12
Sequence 12, Application US/0847733B
Patent No. 5981724
GENERAL INFORMATION:
APPLICANT: ARMITAGE, RICHARD
APPLICANT: FANSLAW, WILLIAM
APPLICANT: SPRIGGS, MELANIE
APPLICANT: SRINIVASAN, SUBHASHINI
APPLICANT: GIBSON, MARYLOU
APPLICANT: MORRIS, ARVIA E.
APPLICANT: MCGREW, JEFFERY
TITLE OF INVENTION: NOVEL CYTOKINE THAT BINDS CD40
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: MS Word for Apple Power Macintosh, version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,733B
FILING DATE: June 07, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/249,189
FILING DATE: May 24, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/969,703
FILING DATE: October 23, 1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/805,723
FILING DATE: December 5, 1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/783,707
FILING DATE: October 25, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2802-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2065870430
TELEFAX: 2065870606
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-733B-12

Query Match 44.6%; Score 1110; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

257 HRRLDKIEDERNLHEDVFVKTIORCNTGERSLSLNCCEIKSOFEGFVNDIMLNKEETK 316
47 HRRLDKIEDERNLHEDVFVKTIORCNTGERSLSLNCCEIKSOFEGFVNDIMLNKEETK 106
317 KENSFENQKDONQIAAHYISEASSKTSVLQMAEKGYTMSNNLVTLENGQLTVKRQ 376

Db 107 KENSFEMOGDQNPQIAHVISEASSKTSVLOMAEKGYTMSNNLVTLNKGQOLTVKRQ 166
QY 377 GLYIYAQVTFCSNREASSQAFIASLCLKSPGFRERILLRANHTSSAKPCGQOSIHLG 436
Db 167 GLYIYAQVTFCSNREASSQAFIASLCLKSPGFRERILLRANHTSSAKPCGQOSIHLG 226
QY 437 GFVELQPGASVFVNTDPSQVSHGTGFTSFGLLKL 471
Db 227 GFVELQPGASVFVNTDPSQVSHGTGFTSFGLLKL 261

RESULT 10
US-08-763-995-2
Sequence 2, Application US/08763995
Patent No. 6017527
GENERAL INFORMATION:
APPLICANT: MARASOVSKY, EUGENE
TITLE OF INVENTION: METHOD OF ACTIVATING DENDRITIC CELLS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Power Macintosh 7200/90
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Macintosh, Version #6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/763,995
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/677,762
FILING DATE: 10 JUL 1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2845-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)367-0430
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-763-995-2

Query Match 44.6%; Score 1110; DB 3; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 257 HRRLDKIDERNLHEDFVFMKTICRCNTGERSLSLNCCEIKSOFEGFVNDIMLNKEETK 316
Db 47 HRRLDKIDERNLHEDFVFMKTICRCNTGERSLSLNCCEIKSOFEGFVNDIMLNKEETK 106
QY 317 KENSFEMOGDQNPQIAHVISEASSKTSVLOMAEKGYTMSNNLVTLNKGQOLTVKRQ 376
Db 107 KENSFEMOGDQNPQIAHVISEASSKTSVLOMAEKGYTMSNNLVTLNKGQOLTVKRQ 166
QY 377 GLYIYAQVTFCSNREASSQAFIASLCLKSPGFRERILLRANHTSSAKPCGQOSIHLG 436
Db 167 GLYIYAQVTFCSNREASSQAFIASLCLKSPGFRERILLRANHTSSAKPCGQOSIHLG 226
QY 437 GFVELQPGASVFVNTDPSQVSHGTGFTSFGLLKL 471

Db 227 GFVELQPGASVFVNTDPSQVSHGTGFTSFGLLKL 261

RESULT 11
US-09-088-913A-12
Sequence 12, Application US/09088913A
Patent No. 6087329
GENERAL INFORMATION:
APPLICANT: ARMITAGE, RICHARD
APPLICANT: FANSLON, WILLIAM
APPLICANT: SPRIGGS, MELANIE
APPLICANT: SRINIVASAN, SUBHASHINI
APPLICANT: GIBSON, MARYLOU
APPLICANT: MORRIS, ARVIA E.
APPLICANT: MCGREW, JEFFERY
TITLE OF INVENTION: NOVEL CYTOKINE THAT BINDS CD40
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: MS Word for Apple Power Macintosh, version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/088,913A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/484,624
FILING DATE:
APPLICATION NUMBER: 08/477,733
FILING DATE: June 07, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/969,703
FILING DATE: October 23, 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/805,723
FILING DATE: December 5, 1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/783,707
FILING DATE: October 25, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2802-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2065870430
TELEFAX: 2065870606
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-088-913A-12

Query Match 44.6%; Score 1110; DB 3; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 257 HRRLDKIDERNLHEDFVFMKTICRCNTGERSLSLNCCEIKSOFEGFVNDIMLNKEETK 316
Db 47 HRRLDKIDERNLHEDFVFMKTICRCNTGERSLSLNCCEIKSOFEGFVNDIMLNKEETK 106
QY 317 KENSFEMOGDQNPQIAHVISEASSKTSVLOMAEKGYTMSNNLVTLNKGQOLTVKRQ 376

Db 107 KENSFEMOKDQNPOLIAHAISEASSTTISVLQAEKGYTMSNNLVTLENGKOLTVKRQ 166
QY 377 GLYIYIAQVTFCSNREASQAAPFIASLCLKSPGRFERILLRAANTHSSARPCGQOSIHLG 436
Db 167 GLYIYIAQVTFCSNREASQAAPFIASLCLKSPGRFERILLRAANTHSSARPCGQOSIHLG 226
QY 437 GFVELQPGASVFVNVTPDPSOVSHGTGTSFGLKL 471
Db 227 GFVELQPGASVFVNVTPDPSOVSHGTGTSFGLKL 261

RESULT 12

US-08-589-771B-8
Sequence 8, Application US/08589771B
Patent No. 6106832
GENERAL INFORMATION:
APPLICANT: ARMITAGE, RICHARD
APPLICANT: DAVISON, BARRY
APPLICANT: FANSLON, WILLIAM
APPLICANT: RENSLOW, BLAIR
APPLICANT: SPRIGGS, MELANIE
APPLICANT: WIDMER, MICHAEL
TITLE OF INVENTION: TREATMENT OF INDIVIDUALS EXHIBITING
TITLE OF INVENTION: DEFECTIVE CD40L (as amended)
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS/Windows 95
SOFTWARE: Word for Windows 95, 7.0a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/589,771B
FILING DATE: January 22, 1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/009,258
FILING DATE: 01/22/93
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HENRY, JANIS C.
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2810-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2065870430
TELEFAX: 2065870606
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-589-771B-8

Query Match 44.6%; Score 1110; DB 3; Length 261;
Best Local Similarity 100.0%; Pred. No. 3, 5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 HRRIDKIEDERNLHEDVFVKTTIQRCTNGERSLSLNCCEIKSQFEGFVKDIMLNKEETK 316
Db 47 HRRIDKIEDERNLHEDVFVKTTIQRCTNGERSLSLNCCEIKSQFEGFVKDIMLNKEETK 106
QY 317 KENSFEMOKDQNPOLIAHAISEASSTTISVLQAEKGYTMSNNLVTLENGKOLTVKRQ 376
Db 107 KENSFEMOKDQNPOLIAHAISEASSTTISVLQAEKGYTMSNNLVTLENGKOLTVKRQ 166

QY 377 GLYIYIAQVTFCSNREASQAAPFIASLCLKSPGRFERILLRAANTHSSARPCGQOSIHLG 436
Db 167 GLYIYIAQVTFCSNREASQAAPFIASLCLKSPGRFERILLRAANTHSSARPCGQOSIHLG 226
QY 437 GFVELQPGASVFVNVTPDPSOVSHGTGTSFGLKL 471
Db 227 GFVELQPGASVFVNVTPDPSOVSHGTGTSFGLKL 261

RESULT 13

PCT-US93-10034-4
Sequence 4, Application PC/TUS9310034
GENERAL INFORMATION:
APPLICANT: Spriggs, Melanie
APPLICANT: Srinivasan, Subhashini
TITLE OF INVENTION: Methods of Preparing Soluble, Oligomeric
TITLE OF INVENTION: Proteins
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10034
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 1003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-10034-4

Query Match 44.6%; Score 1110; DB 5; Length 261;
Best Local Similarity 100.0%; Pred. No. 3, 5e-79;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 HRRIDKIEDERNLHEDVFVKTTIQRCTNGERSLSLNCCEIKSQFEGFVKDIMLNKEETK 316
Db 47 HRRIDKIEDERNLHEDVFVKTTIQRCTNGERSLSLNCCEIKSQFEGFVKDIMLNKEETK 106
QY 317 KENSFEMOKDQNPOLIAHAISEASSTTISVLQAEKGYTMSNNLVTLENGKOLTVKRQ 376
Db 107 KENSFEMOKDQNPOLIAHAISEASSTTISVLQAEKGYTMSNNLVTLENGKOLTVKRQ 166
QY 377 GLYIYIAQVTFCSNREASQAAPFIASLCLKSPGRFERILLRAANTHSSARPCGQOSIHLG 436
Db 167 GLYIYIAQVTFCSNREASQAAPFIASLCLKSPGRFERILLRAANTHSSARPCGQOSIHLG 226
QY 437 GFVELQPGASVFVNVTPDPSOVSHGTGTSFGLKL 471
Db 227 GFVELQPGASVFVNVTPDPSOVSHGTGTSFGLKL 261

RESULT 14

us-09-454-223-2.rai

Page 8

```

1 Patent No. 5961974
2 GENERAL INFORMATION:
3 APPLICANT: ARMITAGE, RICHARD
4 APPLICANT: FANSLAW, WILLIAM
5 APPLICANT: SPRIGGS, MELANIE
6 APPLICANT: SRINIVASAN, SUBHASHINI
7 APPLICANT: GIBSON, MARYLOU
8 TITLE OF INVENTION: NOVEL CYTOKINE
9 NUMBER OF SEQUENCES: 23
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: IMADNEX CORPORATION
12 STREET: 51 UNIVERSITY STREET
13 CITY: SEATTLE
14 STATE: WASHINGTON
15 COUNTRY: USA
16 ZIP: 98101
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: Apple Macintosh
21 OPERATING SYSTEM: Apple Operating System 7.1
22 SOFTWARE: Microsoft Word for Apple, version 5.1a
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/249,189
25 FILING DATE:
26 CLASSIFICATION: 514
27
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: 07/969,703
30 FILING DATE: October 23, 1992
31 CLASSIFICATION: 514
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: 07/805,723
34 FILING DATE: December 5, 1991
35 CLASSIFICATION: 514
36
37 PRIOR APPLICATION DATA:
38 APPLICATION NUMBER: 07/783,707
39 FILING DATE: October 25, 1991
40 CLASSIFICATION: 514
41 ATTORNEY/AGENT INFORMATION:
42 NAME: Perkins, Patricia A.
43 REGISTRATION NUMBER: 34,693
44 REFERENCE/DOCKET NUMBER: 2802-C
45 TELECOMMUNICATION INFORMATION:
46 TELEPHONE: 2065870430
47 TELEFAX: 2065870606
48 INFORMATION FOR SEQ ID NO: 21:
49 SEQUENCE CHARACTERISTICS:
50 LENGTH: 273 amino acids
51 TYPE: amino acid
52 TOPOLOGY: linear
53 MOLECULE TYPE: protein
54
55 US-08-249-189-21

```

| | Query Match | 44.08% | Score 1094.5 | DB 2 | Length 273 |
|----|-----------------------|---|-----------------|---------------|------------|
| | Best Local Similarity | 87.9% | Pred. No. 6e-78 | | |
| | Matches 217 | Conservative | 8 | Mismatches 17 | Indels 5 |
| | | | | Gaps | 2 |
| QY | 226 | LROOMEAKKRLORLEVAESHYKAAALFPDGR-RLDKIEDERNLHEDPFNMKTIONCNT | 284 | | |
| DB | 31 | IEDIEIELSLKIYHINETAIRIKLIL---GERTSSDKIEDERNLHEDPFNMKTIONCNT | 86 | | |
| QY | 285 | GERSLSLNCEIEIKSOFEGFVKDMLNKEERKKEKSEPMQGGDNPPIAAHVISEASSST | 344 | | |
| DB | 87 | GERSLSLNCEIEIKSOFEGFVKDMLNKEERKKEKSEPMQGGDNPPIAAHVISEASSST | 146 | | |
| QY | 345 | TSYLOMAEGGYTMSNNLYLTLENGQLTVKROGLYYTAAQVTEPCSNREASSQAPFIASLC | 404 | | |
| DB | 147 | TSYLOMAEGGYTMSNNLYLTLENGQLTVKROGLYYTAAQVTEPCSNREASSQAPFIASLC | 206 | | |
| QY | 405 | LKSPGFPERILLRAANTHTSSAPCCQOQSILHIGVFEILOPGSVFVNAVTPSPQVSHGTGT | 464 | | |
| DB | 207 | LKSPGFPERILLRAANTHTSSAPCCQOQSILHIGVFEILOPGSVFVNAVTPSPQVSHGTGT | 266 | | |

Tue Jul 3 16:26:04 2001

us-09-454-223-2.ral

Page 9

OY 465 SFGJLKL 471
| | | | |
Db 267 SFGJLKL 273

Search completed: July 3, 2001, 16:15:31
Job time: 108 sec

THIS PAGE BLANK (USPTO)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 08-Oct-1995

C:Accession: I53476; S28017; JH0793; S26694; S28852; S25684; S30593
R:Gauchat, J.
FEBS Lett. 315, 259-266, 1992
A:Title: Human CD40-ligand: Molecular cloning, cellular distribution and regulation of
A:Reference number: I53476; MUID:93138085
A:Accession: I53476
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-261 <RES>
A:Cross-references: GB:I07414; NID:g180123; PIDN:AAA3662.1; PID:g180124
R:Hollenbach, D.; Grosmaire, L.S.; Killis, C.D.; Chalupny, N.J.; Braesch-Andersen, S.;
EMBO J. 11, 4313-4321, 1992
A:Title: The human T cell antigen gp39, a member of the TNF gene family, is a ligand for
A:Reference number: S28017; MUID:93049181
A:Accession: S28017
A:Molecule type: mRNA
A:Residues: 1-261 <HOL>
A:Cross-references: EMBL:Z15017; NID:g38483; PIDN:CAA78737.1; PID:g38484
R:Spiegel, M.K.; Armitage, R.J.; Strohman, L.; Clifford, K.N.; Macduff, B.M.; Sato, T.
J. Exp. Med. 176, 1543-1550, 1992
A:Title: Recombinant human CD40 ligand stimulates B cell proliferation and immunoglobulin
A:Reference number: JH0793; MUID:93094757
A:Accession: JH0793
A:Molecule type: mRNA
A:Residues: 1-261 <SPR>
A:Cross-references: GB:X67878; NID:g38411; PIDN:CAA48077.1; PID:g38412
R:Graf, D.; Korthauer, U.; Mages, H.W.; Senger, G.; Kroczeck, R.A.
Eur. J. Immunol. 22, 3191-3194, 1992
A:Title: Cloning of TRAP, a ligand for CD40 on human T cells.
A:Reference number: S26694; MUID:93076854
A:Accession: S26694
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-261 <GRA>
A:Cross-references: EMBL:X68550; NID:g37269; PIDN:CAA48554.1; PID:g37270
R:Gauchat, J.F.; Aubry, J.P.; Mazzei, G.; Liife, P.; Domotte, T.; Elson, G.; Bonnefoy, J.
FEBS Lett. 315, 259-266, 1993
A:Title: Human CD40-ligand: molecular cloning, cellular distribution and regulation of
A:Reference number: S28852; MUID:93138085
A:Accession: S28852
A:Molecule type: mRNA
A:Residues: 1-261 <RAU>
A:Cross-references: EMBL:L07414; NID:g180123; PIDN:AAA3662.1; PID:g180124
A:Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 6-Gln
C:Genetics:
A:Gene: GDB:CD40LG; HIGM1; IMD3
A:Cross-references: GDB:120632; OMIM:308230
A:Map position: Xq26-Xq26
C:Keywords: glycoprotein; transmembrane protein
F:13-44/Domain: transmembrane #status predicted <TM>
F:45-261/Domain: extracellular #status predicted <EXT>
F:6-240/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 44.6%; Score 1110; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 7, 4e-62;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 257 HNRLLKIEDERLHEDPVMKTIORCNGERSLILNCEELKSGFEGVVKIMLKEETK 316
Db 47 HNRLLKIEDERLHEDPVMKTIORCNGERSLILNCEELKSGFEGVVKIMLKEETK 106
Db 317 KINSEKMGKGDNDPOIAAHVISEASSKTTSVLQNAEKGYTMSNNLVLENGKOLYTKRQ 376
Db 107 KINSEKMGKGDNDPOIAAHVISEASSKTTSVLQNAEKGYTMSNNLVLENGKOLYTKRQ 166
Db 377 GLYYTAAVTCNSNEASSQAPFLASICLKSPGRERILLPAANTHSSAKPCGOOSIHG 436
Db 167 GLYYTAAVTCNSNEASSQAPFLASICLKSPGRERILLPAANTHSSAKPCGOOSIHG 226
Db 437 GVPELPGASVFNVTDPQSVSHGTGFTSGFLKL 471

Db 227 GVPELPGASVFNVTDPQSVSHGTGFTSGFLKL 261
RESULT 3
A45225
Pulmonary surfactant protein D precursor - human
N:Alternate names: SP-D
C:Species: Homo sapiens (man)
C:Date: 16-Apr-1999 #sequence revision 16-Apr-1999 #text change 22-Jun-1999
C:Accession: A45225; S23434; S24555; S44420; S18382; A56776
R:Crouch, E.; Rust, K.; Velle, R.; Donis-Keller, H.; Grosso, L.
J. Biol. Chem. 268, 2976-2983, 1993
A:Title: Genomic organization of human surfactant protein D (SP-D). SP-D is encoded o
A:Reference number: A45225; MUID:93155122
A:Accession: A45225
A:Molecule type: DNA
A:Residues: 1-375 <CRO>
A:Cross-references: GB:L05483; GB:L05484; GB:L05485; NID:g292505; PIDN:AA59450.1; PI
A:Experimental source: Placenta
R:Lu, J.; Willis, A.C.; Reid, K.B.M.
Biochem. J. 284, 795-802, 1992
A:Title: Purification, characterization and cDNA cloning of human lung surfactant pro
A:Reference number: S23434; MUID:92322003
A:Accession: S23434
A:Molecule type: mRNA
A:Residues: 1-30, 'T', 32-121, 'P', 123-179, 'A', 181-375 <LUJ>
A:Cross-references: EMBL:X65018; NID:g34766; PIDN:CAA46152.1; PID:g34767
A:Experimental source: Lung
A:Accession: S24555
A:Molecule type: protein
A:Residues: 214-234, 'X', 236, 'XX', 239-241 <LUJ2>
R:Hoppe, H.J.; Barlow, P.N.; Reid, K.B.M.
FEBS Lett. 344, 191-195, 1994
A:Title: A parallel three stranded alpha-helical bundle at the nucleation site of col
A:Reference number: S44420; MUID:94244769
A:Accession: S44420
A:Molecule type: mRNA
A:Residues: 202-257 <HOP>
R:Rust, K.; Grosso, L.; Zhong, Y.; Chang, D.; Persson, A.; Longmore, W.; Cai, G.Z.; C
Arch. Biochem. Biophys. 290, 116-126, 1991
A:Title: Human surfactant protein D: SP-D contains a C-type lectin carbohydrate recog
A:Reference number: S18382; MUID:91378578
A:Accession: S18382
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 'F', 60-205, 'P', 207-374, 'HF', <RUS>
A:Cross-references: GB:L05485; NID:g292505
A:Note: corrections to this sequence are reported in reference A56776
R:Crouch, E.; Persson, A.; Chang, D.
Am. J. Pathol. 142, 241-248, 1993
A:Title: Accumulation of surfactant protein D in human pulmonary alveolar proteinosis
A:Reference number: A56776; MUID:93142849
A:Accession: A56776
A:Status: preliminary
A:Molecule type: protein
A:Residues: 46-58, 'F', 60-62, 'E', 64-72, 223-227, 'X', 229-239, 'P', 241-245, 'X', 247-256, 'X'
A:Cross-references: PIDN:AA52037.1; PID:g263973; PIDN:AA52038.1; PID:g263974
A:Experimental source: bronchoalveolar lavage
A:Note: sequence extracted from NCBI backbone (NCBI:123024, NCBI:123023); sequence
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lower
C:Genetics: This protein is synthesized by alveolar type II cells.
A:Gene: GDB:SFTPD; SFTPD; SP-D
A:Cross-references: GDB:132674; OMIM:178635
A:Map position: 10q22.2-10q23.1
C:Superfamily: Pulmonary surfactant protein D, C-type lectin homology
F:1-20/Domain: blocked amino end; calcium, glycoprotein; hydroxyllysine; hydroxyproline;
F:21-375/Product: signal sequence #status predicted <SIG>
F:21-45/Domain: non-collagenous #status predicted <MAT>
F:46-222/Domain: collagenous #status predicted <COL>
F:223-375/Domain: non-collagenous #status predicted <NC2>

QY 121 PSKQGNIGPGKPGKGEAGPGMGSTGAKGSGTGPCKGAPGVOGAPGNAG 180
 DB 119 PSKQSGSMGPGTGPFGKGTGPGKGVAPGIDQ--GPPGSGLKGTGKGAAGGAPGAGHAG 175
 QY 181 AAGPAGPAGGAGGAGSGPPGLKGRGVDPDRKIGKESGLPDSALRQOMALKGLQRL 240
 DB 176 VTGSGAIGGQPGSGAGPGLGDKRDPETGAKGESGLAEVNAKQVITLDGHLRF 235
 QY 241 EVAFSHYOKAALFPDGRRLDKIEDERNLHEDFVFMKTQRCNTGERSLSLNCSEIKSQ 300
 DB 236 QNAFSQYKKAVALFPDGOAVGEKI-----FKTAGAVKYSDAEQL-----CREAKG- 280
 QY 301 FEEFVADIMLNKEETKENSEFEMQKQDNPQIAAHVISEASSTKTSVLQMAEKGYTMSN 360
 DB 281 -----QLASPRSSAENAVYQVRAQEKNAIYLSMN 310
 QY 361 NLVT 364
 DB 311 DIST 314

SUIT 7 10450

conglutinin precursor - bovine
 M:Alternate names: C3b-binding protein
 N:Contains: conglutinin-N
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: J00450, J02396, S33235, A23740, S36879, S35044, I46010, A29416, S34054
 R:Suwiki, Y., Yin, Y.P., Makino, M., Kurimura, T., Wakamiya, N.
 Biochem. Biophys. Res. Commun. 191, 335-342, 1993
 A:Title: Cloning and sequencing of a cDNA coding for bovine conglutinin.
 A:Reference number: J00450; MUID:93213261
 A:Accession: J00450
 A:Molecule type: mRNA
 A:Residues: 1-371 <SUZ>
 A:Cross-references: DDBJ:D14085; NID:9285643; PIDN:BA03170.1; PID:9285644
 R:Experimental source: Liver
 R:Kawasaki, N.; Itoh, N.; Kawasaki, T.
 Biochem. Biophys. Res. Commun. 198, 597-604, 1994
 A:Title: Gene organization and 5'-flanking region sequence of conglutinin: A C-type mann
 A:Reference number: J02396; MUID:94128104
 A:Accession: J02396
 A:Molecule type: mRNA
 A:Residues: 1-371 <KA2>
 A:Note: The authors translated the codon GAT for residues 250 and 270 as Glu
 R:Lu, J.; Laursen, S.B.; Thiel, S.; Jensenius, J.C.; Reid, K.B.M.
 Biochem. J. 292, 157-162, 1993
 A:Title: The cDNA cloning of conglutinin and identification of liver as a primary site c
 A:Reference number: S33235; MUID:93277452
 A:Accession: S33235
 A:Molecule type: mRNA
 A:Residues: 1-172, 'H', 174-217, 'A', 219-271, 'V', 273-371 <LU0>
 A:Cross-references: EMBL:X71774; NID:9395267; PIDN:CA05065.1; PID:9395268
 R:Lee, Y.M.; Leiby, K.R.; Allard, J.; Paris, K.; Lerch, B.; Okarma, T.B.
 J. Biol. Chem. 266, 2715-2723, 1991
 A:Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin f
 A:Reference number: A23740; MUID:91131556
 A:Accession: A23740
 A:Molecule type: protein
 A:Residues: 21-209, 'S', 211-371 <LE0>
 R:Kawasaki, N.; Yokota, Y.; Kawasaki, T.
 Arch. Biochem. Biophys. 305, 533-540, 1993
 A:Title: Differentiation of conglutination activity and sugar-binding activity of conglu
 A:Reference number: S36879; MUID:93384312
 A:Accession: S36879
 A:Molecule type: protein
 A:Residues: 21-54, 75-86, 'X', 88-89, 'X', 91, 'X', 93-94, 208-209, 'X', 211-227 <KA0>
 R:Lu, J.; Wiedemann, H.; Holmskov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.
 Eur. J. Biochem. 215, 793-799, 1993
 A:Title: Structural similarity between lung surfactant protein D and conglutinin. Two di

A:Reference number: S35044; MUID:93358905
 A:Accession: S35044
 A:Molecule type: protein
 A:Residues: 75-86, 'X', 88-89, 'X', 91, 'I' <LU0>
 R:Young, N.M.; Leon, M.A.
 Biochem. Biophys. Res. Commun. 143, 645-651, 1987
 A:Title: The carbohydrate specificity of conglutinin and its homology to proteins in
 A:Reference number: A29416; MUID:87184551
 A:Accession: A29416
 A:Contents: annotation
 R:Kawachi, R.; Laursen, S.B.; Willis, A.C.; Slim, R.B.
 Biochem. J. 293, 15-19, 1993
 A:Title: Research Communication. Localization of the receptor-binding site in the col
 A:Reference number: S34054; MUID:93319501
 A:Contents: annotation
 R:Liou, L.S.; Sastiy, R.; Hartsborn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sas
 J. Immunol. 153, 173-180, 1994
 A:Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship
 A:Reference number: I46010; MUID:94267222
 A:Accession: I46010
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-371 <L10>
 A:Cross-references: EMBL:U06860; NID:9507183; PIDN:AB060624.1; PID:9514256
 C:Comment: This protein mediates the agglutination of erythrocytes with antibody and
 C:Comment: This protein is a Ca²⁺-dependent serum lectin specific for N-acetylglucosa
 A:Gene: CGN1
 A:Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1
 C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
 C:Keywords: calcium binding; glycoprotein; hydroxyllysine; hydroxypoline
 F:1-20/Domains: signal sequence #status predicted <Sig>
 F:21-371/Product: conglutinin #status predicted <Sig>
 F:46-214/Region: collagen-like
 F:75-371/Product: conglutinin-N #status predicted <MA2>
 F:248-369/Domains: C-type lectin homology <LCH>
 F:63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #statu
 F:63,87,99,135,141,159,162,198,210/Modified site: 5-hydroxyllysine (Lys) #status exper
 F:78,96,108,111,129,132,147,153,171,195/Modified site: 4-hydroxypoline (Pro) #status
 F:337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 35.3%; Score 878; DB 1; Length 371;
 Best Local Similarity 49.5%; Pred. No. 2, 5e-47;
 Matches 180; Conservative 44; Mismatches 88; Indels 52; Gaps 6;

QY 1 MLPLSLVLLVPLGLGEMSLQSRYPNTCTLWMCSPTEMLGPRDGRGREGPG 60
 DB 3 LPL-LVLLLTLPQWMSLGAEMTTFQKTLANACTLWMCSPLESLGPHGDGREGPHG 61
 QY 61 EKDDPGGLPGMGLSGLOGPTGVPYKPGENGSAEPPKGERGLSGPGLGPGAGREG 120
 DB 62 EKDDPGSGPAGRAGRPGRGWGPIPRGDGVPYGERPKD--TGRPPGPGAGREG 118
 QY 121 PSKQGNIGPGKPGKGEAGPGMGSTGAKGSGTGPCKGAPGVOGAPGNAG 180
 DB 119 PSKQSGSMGPGTGPFGKGTGPGKGVAPGIDQ--GPPGSGLKGTGKGAAGGAPGAGHAG 175
 QY 181 AAGPAGPAGGAGGAGSGPPGLKGRGVDPDRKIGKESGLPDSALRQOMALKGLQRL 240
 DB 176 VTGSGAIGGQPGSGAGPGLGDKRDPETGAKGESGLAEVNAKQVITLDGHLRF 235
 QY 241 EVAFSHYOKAALFPDGRRLDKIEDERNLHEDFVFMKTQRCNTGERSLSLNCSEIKSQ 300
 DB 236 QNAFSQYKKAVALFPDGOAVGEKI-----FKTAGAVKYSDAEQL-----CREAKG- 280
 QY 301 FEEFVADIMLNKEETKENSEFEMQKQDNPQIAAHVISEASSTKTSVLQMAEKGYTMSN 360
 DB 281 -----QLASPRSSAENAVYQVRAQEKNAIYLSMN 310
 QY 361 NLVT 364
 DB 311 DIST 314

RESULT 8

CD40 ligand - mouse
 S21738
 C:Species: Mus musculus (house mouse)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
 C:Accession: S21738
 R:AmiLage, R.J.; Fenslow, W.C.; Strockbine, L.; Sato, T.A.; Clifford, K.N.; Macduff, B.; Cosman, D.; Spriess, M.K.
 Nature 357, 80-82, 1992
 A:Title: Molecular and biological characterization of a murine ligand for CD40.
 A:Reference number: S21738; MID:92244364
 A:Accession: S21738
 A:Molecule type: mRNA
 A:Residues: 1-260 <NR>
 A:Cross-references: EMBL:X65453; NID:950351; PIDN:CAA6448.1; PID:950352
 C:Keywords: glycoprotein; transmembrane protein
 F:23-46/Domain: transmembrane #status predicted <TM>
 F:47-260/Domain: extracellular #status predicted <EXT>
 F:239/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 33.3%; Score 827; DB 2; Length 260;
 Best Local Similarity 64.8%; Pred. No. 2.3e-44;
 Matches 171; Conservative 29; Mismatches 52; Indels 12; Gaps 4;

QY 209 PCDRGIGSGGLPDSALRQOMALKGLQRL-EVAFSHYOKAALFPDGHRLDKIEDER 267
 DB 8 PPSRSV--ATGLPASMKIFMYLLTFLITOMIGSLFAVLT-----HRLDKVEEV 57
 QY 268 NLHEDFEMKTIOKRCNTERSLNCEIKSOEGFVKIDMLNKERTKENSEMQRKD 327
 DB 58 NLHEDFEMKTIOKRCNTERSLNCEIKSOEGFVKIDMLNKERTKENSEMQRKD 116
 QY 328 ONPDIAHVISEASSKTTYSVLQNAEKGYTMSNNLTLENGKOTLVKROGLYTYAOTF 387
 DB 117 EDPDIAHVISEASSKTTYSVLQNAEKGYTMSNNLTLENGKOTLVKROGLYTYAOTF 176
 QY 388 CSNRASSQAPFLASLCLKSPREFRILLNANTHSSAKPGQGSITLGVFELQPCASV 447
 DB 177 CSNRASSQAPFLASLCLKSPREFRILLNANTHSSAKPGQGSITLGVFELQPCASV 236
 QY 448 FVNTDPSOVSHGTGFTSFLGLKL 471
 DB 237 FVNTDPSOVSHGTGFTSFLGLKL 260

RESULT 9

A53570
 collectin-43 - bovine
 N:Alternate names: lectin CL-43
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 12-Apr-1995 #sequence_revision 23-Feb-1996 #text_change 17-Mar-1999
 C:Accession: A53570; A46689
 R:Lim, B.L.; Willis, A.C.; Reid, K.B.M.; Lu, J.; Laursen, S.B.; Jensenius, J.C.; Holmskov, J. Biol. Chem. 269, 11820-11824, 1994
 A:Title: Primary structure of bovine collectin-43 (CL-43). Comparison with conglutinin A
 A:Reference number: A53570; MID:94216283
 A:Accession: A53570
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-301 <LIM>
 A:Cross-references: GB:X75912
 R:Holmskov, J.; Telander, B.; Willis, A.C.; Reid, K.B.; Jensenius, J.C.
 J. Biol. Chem. 268, 10120-10125, 1993
 A:Title: Purification and characterization of a bovine serum lectin (CL-43) with structure
 A:Reference number: A46689; MID:93252891
 A:Accession: A46689
 A:Molecule type: protein
 A:Residues: 1-27 <HOL>
 A:Experimental source: serum
 A:Note: sequence extracted from NCBI backbone (NCBIP:J31234)

C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
 C:Keywords: lectin
 F:177-299/Domain: C-type lectin homology <LCH>

Query Match 22.7%; Score 565.5; DB 2; Length 301;
 Best Local Similarity 35.8%; Pred. No. 4.1e-28;
 Matches 127; Conservative 42; Mismatches 81; Indels 105; Gaps 8;

QY 21 EMKSLQSRVPMCTLYVSGPTENGSLPDRDGRDREGPRGKGGPGLPGMGLSGLOGPT 80
 DB 2 EMDVSEKTLTDPCITLVACAPADSLRGHDGRDREGPRGKGGP----- 46
 QY 81 GPVPRKENGNSAGPRGKGRGLSGPPGLGIPDPARKDEPSGKQGNIGPGKPRGGA 140
 DB 47 -----GPPGM---PGPAREGSGKQSGMGPPTGPRGP 79
 QY 141 GPKGEVAGPMQSGTGAKSGTPKGERGAPGVQAGPAGPAGPAGPAGPAGPAGP 200
 DB 80 GPEGVAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 124
 QY 201 GLKGRGVPGDRGIGSGGLPDSALRQOMALKGLQRL-EVAFSHYOKAALFPDGHRL 260
 DB 125 GLKGRGVPGDRGIGSGGLPDSALRQOMALKGLQRL-EVAFSHYOKAALFPDGHRL 184
 QY 261 DKIDERNLHEDFEMKTIOKRCNTERSLNCEIKSQF-----EGF 304
 DB 185 EK1-----FKTAGAVSYSDAEQL-----CREAKGGLASSPSSEANEAVTLVRAK 230
 QY 305 VKDMLNKERTKENSEMQRKD-----DONPDIAHVISEAS 341
 DB 231 NKHAHVISEASSKTTYSVLQNAEKGYTMSNNLTLENGKOTLVKROGLYTYAOTF 285

RESULT 10

S28791
 collagen alpha 1(XI) chain - chicken (fragment)
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
 C:Accession: S28791
 R:Nah, H.D.; Barembaum, M.; Upholt, W.B.
 J. Biol. Chem. 267, 22581-22586, 1992
 A:Title: The chicken alpha1(XI) collagen gene is widely expressed in embryonic tissue
 A:Reference number: S28791; MID:93054557
 A:Accession: S28791
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-888 <NAH>
 A:Cross-references: EMBL:M88593; NID:9211619; PIDN:AAA48707.1; PID:9211620
 C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homolo
 F:665-887/Domain: fibrillar collagen carboxyl-terminal homology <FCO>

Query Match 19.6%; Score 488.5; DB 2; Length 888;
 Best Local Similarity 47.7%; Pred. No. 8.3e-23;
 Matches 102; Conservative 17; Mismatches 62; Indels 33; Gaps 4;

QY 41 PTENGLPGRDGRDREGPRGKGGPGLPGMGLSGLOGPTGPVGP-----GE 88
 DB 399 PCPPGEPGPAGQDGVGEGEDPDGPGCPGPGSGAGPDPGPKRPPATGAEGRGE 458
 QY 89 NSAGEP-----GPKRGGLSGPPGLPGITGPAGK-----GPGKQGNIGP 130
 DB 459 KGAGGEGAGCAPKTKTPVGPQGPAGPAGPGLGIGIPVGEDGLPGAPQDGPGLGP 518
 QY 131 QGRGPGGEAGPKEVAGPAGMGTGAKSGTPKGERGAPGVQAGPAGPAGPAGPAGP 187
 DB 519 PGLPGLKGDGSGKEKHPGLIGLIGPREGDEKGDGLPGQGSPPAKKDAGISGAPG 578
 QY 188 AGPGAPGSRGPPGLKGRGVPGDRGIGSGGLP 221
 DB 579 LGPPGPGLPQGPQGPAGKSSSGPAGQKQDGLP 612

RESULT 11

CGBO7S
 Collagen alpha 1(III) chain - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 04-Dec-1986 #sequence, revision 04-Dec-1986 #text, change 07-May-1999
 C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946
 R:Filetek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wächter, E.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
 A:Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen
 A:Accession: A02862; MUID:80026026
 A:Molecule type: protein
 A:Residues: 1-242 <EXP>
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
 A:Title: The covalent structure of calf skin type III collagen. II. The amino acid sequen
 A:Accession: A38001; MUID:80026027
 A:Molecule type: protein
 A:Residues: 243-422 <DEW>
 R:Bentz, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
 A:Title: The covalent structure of calf skin type III collagen. III. The amino acid sequ
 A:Accession: A38002; MUID:80026028
 A:Molecule type: protein
 A:Residues: 423-571 <DEW>
 R:Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
 A:Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequ
 A:Accession: A38003; MUID:80026029
 A:Molecule type: protein
 A:Residues: 572-808 <LAN>
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
 A:Title: The covalent structure of calf skin type III collagen. V. The amino acid sequen
 A:Accession: A38004; MUID:80026030
 A:Molecule type: protein
 A:Residues: 809-947 <DEW2>
 R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
 A:Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequen
 A:Accession: A38005; MUID:80026031
 A:Molecule type: protein
 A:Residues: 948-1049 <ALL>
 R:Henkel, W.
 Biochem. J. 318, 497-503, 1996
 A:Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
 A:Accession: S71946; MUID:96404897
 A:Molecule type: protein
 A:Residues: 87-106;1017-1029;1037-1049 <HBN>
 C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are h
 C:Superfamily: The type III collagen molecule is a trimer of identical chains, linked to eac
 C:Keywords: coiled coil; extracellular matrix; fibrillar collagen carboxyl-terminal homology;
 F:1-1049/Product: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 F:1-14/Region: amino-terminal nonhelical telopeptide
 F:15-1040/Region: helical
 F:587-589/Region: cell attachment (R-G-D) motif
 F:752-754/Region: cell attachment (R-G-D) motif
 F:875-877/Region: cell attachment (R-G-D) motif
 F:878-880/Region: cell attachment (R-G-D) motif
 F:935-937/Region: cell attachment (R-G-D) motif
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide
 F:953-107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental

F:1040,1041/Disulfide bonds: Interchain #status predicted

Query Match 19.5%; Score 484.5; DB 1; Length 1042;
 Best Local Similarity 41.2%; Pred. No. 1,8e-22;
 Matches 113; Conservative 13; Mismatches 71; Indels 77; Gaps 9;
 14 PLGLTGLKEMKSLQSRVPTCTLWMSPTENGLPGDRGRDRE-----GPRGKRGDGL 67
 Db 427 PKGNDGAPCKN-GERGPG-----GPGPGPAGKNGETGPGPGTSGDNGDTP 478
 68 PGPMGLSLGQPTGPGVPGKENGSGAGEPGKRGRL-----SGPGLRGI 112
 Db 479 PGPGQLGLPTGS---GPGENKRRPEPCKGEACAPICPGKDGSGAGRGCPGAGCP 535
 113 PGAGAGEGSSGKGNIGPOGKFGPKGEA-----GPKGEVAPGMQGS 154
 Db 536 PGPAGGAPGPGPGCGKAAGPPGPGSAGTPGLQGMPERGPGPGPKGDKGPGSSGV 595
 155 TGAKGSTGP-----KGERGARGV-----QGAPGNAGAPGAPGAP 190
 Db 596 DGAPGKIDGRPTGPRPGAPGQPDGSGSAGVPGIACPGPBERGPGCPGAGF 655
 191 QGAPGSRPGPLKGRGVPGDRGIRGESGLPDSA 224
 Db 656 PGPQNGEPGAKGERGARGE---KGEGGPGMA 686

RESULT 12

CGCH1S

Collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
 C:Species: Gallus gallus (chicken)
 C:Date: 12-Aug-1981 #sequence, revision 06-Jul-1982 #text, change 31-Mar-2000
 C:Accession: A90458; A90181; A02857
 R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross,
 Biochemistry 21, 2048-2055, 1982
 A:Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete pr
 A:Accession: A90458; MUID:82231995
 A:Molecule type: protein
 A:Residues: 1-1036 <HIG>
 A:Experimental source: skin
 A:Note: This is the latest in a series of papers from these workers elucidating the s
 R:Kjell  , D.R.; Glimcher, M.J.
 Biochem. Biophys. Res. Commun. 48, 720-726, 1972
 A:Title: Evidence for a previously undetected sequence at the carboxyterminus of the
 A:Accession: A90181; MUID:72243016
 A:Molecule type: protein
 A:Residues: 1037-1042 <EXP>
 A:Experimental source: skin
 A:Note: residues 1037-1042 above correspond to the carboxyl end of the protein
 C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in s
 C:Comment: Most of the prolines at the third position of the tripeptide repeating uni
 C:Superfamily: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trime
 F:1/Modified site: pyroglutamic acid (Gln) #status experimental

Query Match 19.3%; Score 481; DB 1; Length 1042;
 Best Local Similarity 51.9%; Pred. No. 2.9e-22;
 Matches 97; Conservative 9; Mismatches 75; Indels 6; Gaps 2;

41 PLENLPGRGDRGREGRGKDPGLPGMLSLGQPTGPGVPRKENGSGAGEPGKGE 100
 Db 367 PDEALPLPAKGLTGSPGSPGDPDKTGPPGAGQGRGCPAPCAKQDAQVWPFPGKGA 426
 101 RGLSPGGLPGIPGAGKESGSGKGNIGPOGKFGPKGEAGVPGAGAPGMQGSTGAKS 160
 Db 427 AGPRKPEERKAPGPAGVGAAGRDGAAGQPGPGPAGGERGEGQAGAPGQGLPGR 486
 161 TGPGERGAPGVGAPGNAGAPGAPGAGPQAGAPGSR---GPGGLAGDR---GVPGDRGI 214

DB 487 AGPGEAGKGECEGVNAGAPGAPGARGERGVGPGPQGPGRGANGAPGNDGA 546
 QY 215 KGESGLP 221
 DB 547 KDGAGAP 553

RESULT 13

CGHUIE

collagen alpha 1(XI) chain precursor - human

N:Alternate names: procollagen alpha 1(XI) chain

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1990 #sequence_revision 03-Oct-1995 #text_change 08-May-1998

C:Accession: A35239; A31795

R:Yoshioke, H.; Ramirez, F.

J. Biol. Chem. 265, 6423-6426, 1990

A:Title: Pro-alpha1(XI) collagen. Structure of the amino-terminal propeptide and expres

A:Reference number: A35239; MUID:90202924

A:Accession: A35239

A:Molecule type: mRNA

A:Residues: 1-558 <YOS>

A:Cross-references: GB:J05407

R:Bernard, M.; Yoshioke, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.;

J. Biol. Chem. 263, 17159-17166, 1988

A:Title: Cloning and sequencing of pro-alpha1(XI) collagen cDNA demonstrates that type X

cartilagenous tissue.

A:Reference number: A92689; MUID:89034222

A:Accession: A31795

A:Molecule type: DNA; mRNA

A:Residues: 538-1806 <BER>

A:Cross-references: GB:J04177

A:Gene: GDB:COL1A1; COL16

A:Map position: 1p21-1p21 597/3; 615/3; 633/3; 648/3; 666/3; 681/3

A:Introns: 561/3; 579/3; 597/3; 615/3; 633/3; 648/3; 666/3; 681/3

A>Note: the list of introns is incomplete

C:Complex: type XI collagen is a heterotrimer of two alpha 1(XI) chains and one alpha

3(XI) chain (see PIR:CGHUC), initially linked by disulfide bonds among their carboxyl-

term with desmosine cross-links made from lysine and allysine residues

C:Function: A:Description: structural component of extracellular fibrous polymer associated with cell

A>Note: may play a role in controlling the lateral growth of collagen II fibrils

C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology

C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli

F:1-36/Domain: signal sequence #status predicted <SIG>

F:35-260/Domain: PARP-like #status predicted <PARP>

F:37-311/Domain: amino-terminal propeptide #status predicted <PRO>

F:512-1565/Product: collagen alpha 1(XI) chain #status predicted <MAT>

F:512-327/Region: amino-terminal nonhelical telopeptide

F:528-1542/Region: helical

F:143-1565/Region: carboxyl-terminal nonhelical telopeptide

F:1565-1806/Domain: carboxyl-terminal propeptide #status predicted <CRP>

F:1583-1805/Domain: fibrillar collagen carboxyl-terminal homology <RCC>

F:61-243,182-236/Disulfide bonds: #status predicted

F:505/Modified site: allysine (lys) #status predicted

F:612,1452/Modified site: 5-hydroxylysine (lys) #status predicted

F:612,1452/Binding site: carbohydrate (lys) (covalent) #status predicted

* Query Match 19.38; Score 480; DB 1; Length 1806;

Best Local Similarity 45.58; Pred. No. 6,4e-22;

Matches 100; Conservative 18; Mismatches 60; Indels 42; Gaps 4;

QY 41 PTEGLGR---DGRDGEGRGKGPGLPGPGLSLGSLGTPP-----82

DB 1314 PGDGPPEELPGACDGVGGGKGGGDDGQPCPPGPPSEADPPGPPGKRGPPGAAGR 1373

QY 83 -----VGPKGENGSAGEPGRGRLSGPGLPGIPGAPGKSPSK 124

DB 1374 OGEKAGKGEAGBEPGKTPGVPGPQGPAGKRGPEGLRGIPGVGEQGLPGAAAGDGPGR 1433
 QY 125 QGNIGPGKGPQKGEAGPKEBVGAPMGSTGAKGTGPGKRGAPGVQCAP---GAGA 181
 DB 1434 ---MGPPGLPGLKDPGSKGEKHPGLGLIGPGEQGEKGRDLPGTQSPGAKDGCI 1490
 QY 182 AGPAGPAGPGAPGSRGPPGLKGRGVPCDRGIRKGESGLP 221
 DB 1491 PGAPGAPPGPPGLPGPQGPAGKRGSTGAPAGKDGSLP 1530

RESULT 14

CGHUIV

collagen alpha 1(V) chain precursor - human

N:Alternate names: procollagen alpha 1(V) chain

C:Species: Homo sapiens (man)

C>Date: 22-Nov-1993 #sequence_revision 03-Oct-1995 #text_change 16-Jun-2000

C:Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665

R:Greenspan, D.S.; Cheng, W.; Hoffman, G.G.

J. Biol. Chem. 266, 24727-24733, 1991

A:Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution o

A:Reference number: S18802; MUID:92105142

A:Accession: S18802

A:Molecule type: mRNA

A:Residues: 1-1838 <GRE>

A:Cross-references: GB:M76729; NID:G189519; PIDN:AAA5993.1; PID:G189520

R:Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaol, Y.; Kato, I.

J. Biol. Chem. 266, 13124-13129, 1991

A:Title: Complete primary structure of human collagen alpha-1(V) chain.

A:Reference number: S16024; MUID:91302336

A:Accession: S16024

A:Molecule type: mRNA

A:Residues: 1-81, 'QL', 84-389, 'A', 391-676, 'K', 678-1294, 'PS', 1297, 'RS', 1300-1553, 'R', 15

A:Cross-references: GB:D90279; NID:G219509; PIDN:BAAL323.1; PID:G219510

A>Note: parts of this sequence were determined by protein sequencing

R:Yaol, Y.; Hashimoto, K.; Takahara, K.; Kato, I.

Exp. Cell Res. 194, 180-185, 1991

A:Title: Insulin binds to type V collagen with retention of mitogenic activity.

A:Reference number: A61142; MUID:91224163

A:Accession: A61142

A:Molecule type: protein

A:Residues: 823-824, 'X', 826-842 <YAO>

A>Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyano

R:Yaol, Y.; Hashimoto, K.; Koltabashi, H.; Takahara, K.; Ito, M.; Kato, I.

Biochim. Biophys. Acta 1035, 139-145, 1990

A:Title: Primary structure of the heparin-binding site of type V collagen.

A:Reference number: S11303; MUID:90366601

A:Accession: S11303

A:Molecule type: protein

A:Residues: 823-824, 'X', 826-848, 'T', 850-851, 'P', 853, 'PR', 856-893, 'D', 895-932, 'X', 934-

A>Note: the residues designated 'X' are probably glycosylated hydroxylysine; this seq

R:Seyer, J.M.; Kang, A.H.

Arch. Biochem. Biophys. 271, 120-129, 1989

A:Title: Covalent structure of collagen: amino acid sequence of three cyanogen bromid

A:Reference number: S03978; MUID:89227189

A:Accession: S03978

A:Molecule type: protein

A:Residues: 621-640, 'G', 642-649, 'L', 651-662, 'E', 664-667, 'Q', 669-676, 'Q', 678-683, 'P', 6

A>Note: there are a number of inconsistencies between the sequences in figures 6 and

R:Morad-Ameli, M.; Rousseau, J.C.; Klemen, J.P.; Champilaud, M.F.; Bouillon, M.M.;

Eur. J. Biochem. 221, 987-995, 1994

A:Title: Diversity in the processing events at the N-terminus of type-V collagen.

A:Reference number: S43642; MUID:94237164

A:Accession: S43642

A:Molecule type: protein

A:Residues: 565-576, 756-758, 'X', 760-763, 'X', 765-772, 1012-1029, 1219-1232, 1465-1474, 'X'

R:Fessler, L.I.; Brosh, S.; Chaplin, S.; Fessler, J.H.

J. Biol. Chem. 261, 5034-5040, 1986

A:Title: Tyrosine sulfation in precursors of collagen V.

A:Reference number: A56977; MUID:65188226

A:Contents: annotation; identification of tyrosine sulfate in the amino-terminal prop

R:Lee, S.: Greenspan, D.S.

Biochem. J. 310, 15-22, 1995

A:Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).

A:Reference number: S58665; MIMID:95374437

A:Accession: S58665

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-36 <L6E>

C:Cross-references: GB:L38608; NID:91020325; PIDN:AAA79653.1; PID:g1020326

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit are 5-hydroxylated and subsequently O-glycosylated.

C:Comment: A long form of the mature protein containing part of the amino-terminal propeptide the heterotrimers are probably processed to the long form.

C:Genetics:

A:Gene: GDB:COL5A1

A:Cross-references: GDB:131457; OMIM:120215

A:Map position: 9q34.2-9q34.3

C:Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer of alpha 2(V) chain and one alpha 3(V) chain. Initially linked by disulfide bonds among the

C:Function: length, is formed with desmosine cross-links made from lysine and allysine residues

Description: structural component of extracellular fibrous polymer associated with cell

C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology

C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline

F:1-37/Domain: signal sequence #status predicted <SIG>

F:36-261/Domain: PARP-like #status predicted <PARP>

F:38-541/Domain: amino-terminal propeptide #status predicted <PRO>

F:542-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>

F:559-1572/Region: amino-terminal nonhelical telopeptide

F:645-647/Region: helical

F:663-665/Region: cell attachment (R-G-D) motif

F:897-929/Region: cell attachment (R-G-D) motif

F:1573-1605/Region: carboxyl-terminal nonhelical telopeptide

F:1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>

F:1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

F:38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F:155-116,185,1672,1741/Binding site: carboxylate (Asn) (covariant) #status predicted

F:234,236,440,262,263,273,274,275,277,279,280,338,340,346,347,352,357,416,417,420,421/Binding site: allysine (Lys) #status predicted

F:542/Modified site: Ala-Gln (procollagen N-endopeptidase) #status predicted

F:570,576,621,639,648,654,657,678,690,693,696,705,717,720,726,732,741,750,753,756,762/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F:627,642,687,708,744,774,795,804,807,810,819,825,846,864,882,897,1482/Modified site: 5-hydroxylysine (Pro) #status experimental

F:708,744/Binding site: carboxylate (Lys) (covariant) #status experimental

F:1482/Modified site: 5-hydroxylysine (Lys) #status predicted

F:1605-1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted

F:1639,1645,1662,1671/Disulfide bonds: interchain #status predicted

F:1680-1835,1746-1789/Disulfide bonds: #status predicted

Query Match 19.3% Score 479.5; DB 1: Length 1838;
Best Local Similarity 44.1% Pred No. 7e-22;
Matches 105; Conservative 9; Mismatches 67; Indels 57; Gaps 4;

```

QY 41 PRNGLPGDGRGPRGKGGPGLPGMGLSLGPTGVPKGENSGAGEGPKGE 100
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1323 PCPKGPGDGGKRGKSPGVPFPGDPGPPGAGQDGPDDKDDDPGQTSRPGTGE 1382
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 RGLSGPPGLPGIPGPKG-----EGPSGKGNIGPQ---GKGP----- 136
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1383 PCPSGPPKRGPPGAPGPGGKGGKAGKAGLEGGPKGTGPIGPGAPKPGDGLRGI 1442
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 137 -----KGEAGPKGEVAPGMOSTGAKSGTGPKGE 166
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1443 PCPVGEGLPGSPGPPGPPGPPGLGLKGDGSPKGEKCHPGLILGPPGEQGEKGD 1502
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 RGAPGVQGA---PGNAGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 221
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 1503 RGLPGQSSGPKBEGGIGTGFSGPIGPPGPPGLPGPPGPKGAKGSSQPTGPKGEAGHP 1560

RESULT 15

S18803

collagen alpha 1(V) chain - hamster

C:Species: Cricetinae gen. sp. (hamster)

C>Date: 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change 16-Dec-1998

C:Accession: S18803

R:Greenspan, D.S.; Cheng, W.; Hoffman, G.G.

J. Biol. Chem. 266, 24727-24733, 1991

A:Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution o

A:Reference number: S18802; MIMID:92105142

A:Accession: S18803

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1843 <GRE>

C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homolo

F:1620-1842/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 19.3% Score 479.5; DB 2: Length 1843;
Best Local Similarity 44.1% Pred No. 7e-22;
Matches 105; Conservative 9; Mismatches 67; Indels 57; Gaps 4;

```

QY 41 PRNGLPGDGRGPRGKGGPGLPGMGLSLGPTGVPKGENSGAGEGPKGE 100
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1328 PCPKGPGDGGKRGKSPGVPFPGDPGPPGAGQDGPDDKDDDPGQTSRPGTGE 1387
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 RGLSGPPGLPGIPGPKG-----EGPSGKGNIGPQ---GKGP----- 136
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1388 PCPSGPPKRGPPGAPGPGGKGGKAGKAGLEGGPKGTGPIGPGAPKPGDGLRGI 1447
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 137 -----KGEAGPKGEVAPGMOSTGAKSGTGPKGE 166
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1448 PCPVGEGLPGSPGPPGPPGPPGLGLKGDGSPKGEKCHPGLILGPPGEQGEKGD 1507
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 RGAPGVQGA---PGNAGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 221
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1508 RGLPGQSSGPKBEGGIGTGFSGPIGPPGPPGLPGPPGPKGAKGSSQPTGPKGEAGHP 1565
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: July 3, 2001, 16:16:42
Job time: 159 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:15:03 ; Search time 12.03 Seconds
(without alignments)
1341.175 Million cell updates/sec

Title: US-09-454-223-2
Sequence: 1 MRPFLSMVLVLQVPLGNLGA.....TDPQVSHGTGFTSFGCLKL 471
Perfect score: 2487

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Length | ID | Description |
|------------|--------|--------------|--------|-------------|
| 1 | 1384.5 | 55.7 | 374 1 | PSPD_MOUSE |
| 2 | 1286 | 51.7 | 374 1 | PSPD_MOUSE |
| 3 | 1110 | 44.6 | 261 1 | TNFS_HUMAN |
| 4 | 1053 | 42.3 | 375 1 | PSPD_MOUSE |
| 5 | 974 | 39.2 | 261 1 | TNFS_BOVIN |
| 6 | 967.5 | 38.8 | 369 1 | PSPD_MOUSE |
| 7 | 939.5 | 37.8 | 260 1 | TNFS_BOVIN |
| 8 | 907.5 | 36.5 | 260 1 | TNFS_BOVIN |
| 9 | 878 | 35.3 | 371 1 | TNFS_BOVIN |
| 10 | 827 | 33.3 | 260 1 | TNFS_MOUSE |
| 11 | 565.5 | 22.7 | 301 1 | CL43_BOVIN |
| 12 | 484.5 | 19.5 | 1049 1 | CA13_BOVIN |
| 13 | 482.5 | 19.4 | 1460 1 | CA11_CANFA |
| 14 | 481.5 | 19.4 | 1464 1 | CA11_HUMAN |
| 15 | 481 | 19.3 | 1453 1 | CA11_CHICK |
| 16 | 480.5 | 19.3 | 1804 1 | CA1B_MOUSE |
| 17 | 480 | 19.3 | 1806 1 | CA1B_HUMAN |
| 18 | 478 | 19.2 | 1650 1 | CA2B_MOUSE |
| 19 | 477.5 | 19.2 | 1464 1 | CA13_MOUSE |
| 20 | 476.5 | 19.2 | 1453 1 | CA15_HUMAN |
| 21 | 476.5 | 19.2 | 1838 1 | CA15_HUMAN |
| 22 | 475 | 19.1 | 671 1 | CA11_RAT |
| 23 | 474.5 | 19.1 | 1466 1 | CA13_HUMAN |
| 24 | 474 | 19.1 | 1736 1 | CA2B_HUMAN |
| 25 | 473.5 | 19.0 | 1418 1 | CA11_HUMAN |
| 26 | 472.5 | 19.0 | 1779 1 | CA11_BOVIN |
| 27 | 472.5 | 19.0 | 1763 1 | CA24_ASCSU |
| 28 | 470.5 | 18.9 | 911 1 | CA1B_BOVIN |
| 29 | 470 | 18.9 | 1022 1 | CA26_CHICK |
| 30 | 469.5 | 18.9 | 1459 1 | CA12_MOUSE |
| 31 | 469.5 | 18.9 | 2944 1 | CA17_HUMAN |
| 32 | 468.5 | 18.8 | 482 1 | CA1B_BOVIN |
| 33 | 462.5 | 18.6 | 747 1 | CA12_BOVIN |

ALIGNMENTS

| RESULT | ID | STANDARD | PRT | 374 AA. |
|--------|--|----------|-----|---------|
| 1 | PSPD_MOUSE | | | |
| 2 | PSPD_MOUSE | | | |
| 3 | P50404 | | | |
| 4 | 01-OCT-1996 (Rel. 34, Created) | | | |
| 5 | 01-OCT-1996 (Rel. 34, Last sequence update) | | | |
| 6 | 15-JUL-1999 (Rel. 38, Last annotation update) | | | |
| 7 | PULMONARY SURFACTANT-ASSOCIATED PROTEIN D PRECURSOR (SP-D) (PSP-D). | | | |
| 8 | SFTPD OR SFTPA. | | | |
| 9 | Mus musculus (Mouse). | | | |
| 10 | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | |
| 11 | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| 12 | NCBI_Taxid=10090; | | | |
| 13 | [1] | | | |
| 14 | SEQUENCE FROM N. A. | | | |
| 15 | STRATIN-B6/CBA; TISSUE=Lung; | | | |
| 16 | MEDLINE=96094460; Pubmed=7499852; | | | |
| 17 | RA Motwani M., White R.A., Guo N., Dowler L.L., Tauber A.I., Sastry K.N.; | | | |
| 18 | "Mouse surfactant protein-D. cDNA cloning, characterization, and gene | | | |
| 19 | localization to chromosome 14." | | | |
| 20 | J. Immunol. 155:5671-5677(1995). | | | |
| 21 | CC -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED | | | |
| 22 | MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER | | | |
| 23 | EXTENT OTHER ALPHA-GLUCOSYL MOTIFETS. IT COULD PARTICIPATE IN THE | | | |
| 24 | EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT. | | | |
| 25 | CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS. | | | |
| 26 | CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR. | | | |
| 27 | CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% | | | |
| 28 | PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, | | | |
| 29 | CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-B AND SP-C). | | | |
| 30 | CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN. | | | |
| 31 | CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN. | | | |
| 32 | ----- | | | |
| 33 | THIS SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| 34 | between the Swiss Institute of Bioinformatics and the EMBL Outstation - | | | |
| 35 | the European Bioinformatics Institute. There are no restrictions on its | | | |
| 36 | use by non-profit institutions as long as its content is in no way | | | |
| 37 | modified and this statement is not removed. Usage by and for commercial | | | |
| 38 | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| 39 | or send an email to license@isb-sib.ch). | | | |
| 40 | ----- | | | |
| 41 | EMBL; LA0156; AAA92021.1; - | | | |
| 42 | HSSP; P19999; 1MSB. | | | |
| 43 | MGI; MGI:109515; sftpd. | | | |
| 44 | InterPro: IPR000087; - | | | |
| 45 | InterPro: IPR001304; - | | | |
| 46 | Pfam; PF01391; Collagen; 3. | | | |
| 47 | PROSITE; PS00615; C-TYPE LECTIN 1; 1. | | | |
| 48 | PROSITE; PS50041; C-TYPE LECTIN 2; 1. | | | |
| 49 | Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation; | | | |
| 50 | Signal; Lectin; Collagen; Repeat; Coiled coil. | | | |
| 51 | FT SIGNAL 1 19 | | | |
| 52 | FT CHAIN 20 374 | | | |
| 53 | D. | | | |

Query Match
Best Local Similarity 55.7%; Score 1384.5; DB 1; Length 374;
Matches 262; Conservative 0; Mismatches 10; Indels 3; Gaps 1;

FT DOMAIN 45 221 COLLAGEN-LIKE.
FT DOMAIN 222 253 COILED COIL (POTENTIAL).
FT DOMAIN 278 374 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 280 372 BY SIMILARITY.
FT DISULFID 350 364 BY SIMILARITY.
FT CARBOHYD 89 89 N-LINKED (GLCNAC...) (POTENTIAL).
SO SEQUENCE 374 AA; 37688 MW; PE03426126P43E4 CRC64;

QY 1 MLPLSLVLLVPLNLGAEMKLSQSRVPTCTLVMSPTENGLPGDGDGREGPRG 60
DB 1 MLPLSLVLLVPLNLGAEMKLSQSRVPTCTLVMSPTENGLPGDGDGREGPRG 60
QY 61 EKGDPLPGPMGLSGIAGPTGPGKGEVGAAPGAGPAGPAGPAGPAGPAGPAG 120
DB 61 EKGDPLPGPMGLSGIAGPTGPGKGEVGAAPGAGPAGPAGPAGPAGPAGPAG 120
QY 121 PSKGQGNIGPQKPGPKGAGPKGEVGAAPGAGPAGPAGPAGPAGPAGPAG 180
DB 121 PSKGQGNIGPQKPGPKGAGPKGEVGAAPGAGPAGPAGPAGPAGPAGPAG 180
QY 181 AAGPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 240
DB 181 AAGPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 240
QY 241 EVAFSHYOKAALFPDGRHRLDKI--EDERNLHED 272
DB 241 EVAFSHYOKAALFPDGRHRLDKI--EDERNLHED 272

RESULT 2
PSPD_RAT STANDARD; PRT; 374 AA.
AC P35248;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN D PRECURSOR (SP-D) (PSP-D)
GN (CP4).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathia; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-33.
RA MEDLINE=92112913; PubMed=1370483;
RA Shimizu H., Fisher J.H., Papst P., Benson B., Lau K., Mason R.J.,
Voelker D.R.;
RT "Primary structure of rat pulmonary surfactant protein D. cDNA and
RT deduced amino acid sequence.";
RL J. Biol. Chem. 267:1853-1857(1992).
RN [2]
RP SEQUENCE OF 73-95 AND 153-180.
RC TISSUE=Lung;
RX MEDLINE=9001186; PubMed=2675969;
RA Person A., Chang D., Rust K., Moxley M., Longmore W., Crouch E.;
RT "Purification and biochemical characterization of CP4 (SP-D), a
RT collagenous surfactant-associated protein.";
RT Biochemistry 28:6361-6367(1989).
CC -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
CC EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE
CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT.
CC -1- PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN. 2 COLLAGENOUS,

CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).
CC EMBL: M81231; AAA2170.1; -
DR PIR: A43046; A43046.
DR HSSP: P19999; 1MSB.
DR InterPro: IPR000087; -
DR InterPro: IPR001304; -
DR Pfam: PF01391; Collagen; 3.
DR Pfam: PF00059; lectin-c; 1.
DR PROSITE: PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat; Coiled coil.
FT CHAIN 1 19 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT 20 374 D.
FT 374 374 COLLAGEN-LIKE.
FT DOMAIN 45 221 COILED COIL (POTENTIAL).
FT DOMAIN 222 253 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 278 372 BY SIMILARITY.
FT DISULFID 350 364 BY SIMILARITY.
FT CARBOHYD 89 89 N-LINKED (GLCNAC...).
FT MOD_RES 77 77 HYDROXYLATION.
FT MOD_RES 86 86 HYDROXYLATION.
FT MOD_RES 95 95 HYDROXYLATION.
FT MOD_RES 98 98 HYDROXYLATION.
FT MOD_RES 170 170 HYDROXYLATION.
FT MOD_RES 176 176 HYDROXYLATION.
FT CONFLICT 89 89 N->E (IN REF. 2).
FT CONFLICT 164 164 K->C (IN REF. 2).
SO SEQUENCE 374 AA; 37561 MW; DB2B5E399DBA3C CRC64;

Query Match
Best Local Similarity 51.7%; Score 1286; DB 1; Length 374;
Matches 245; Conservative 7; Mismatches 34; Indels 0; Gaps 0;

QY 1 MLPLSLVLLVPLNLGAEMKLSQSRVPTCTLVMSPTENGLPGDGDGREGPRG 60
DB 1 MLPLSLVLLVPLNLGAEMKLSQSRVPTCTLVMSPTENGLPGDGDGREGPRG 60
QY 61 EKGDPLPGPMGLSGIAGPTGPGKGEVGAAPGAGPAGPAGPAGPAGPAGPAG 120
DB 61 EKGDPLPGPMGLSGIAGPTGPGKGEVGAAPGAGPAGPAGPAGPAGPAGPAG 120
QY 121 PSKGQGNIGPQKPGPKGAGPKGEVGAAPGAGPAGPAGPAGPAGPAGPAG 180
DB 121 PSKGQGNIGPQKPGPKGAGPKGEVGAAPGAGPAGPAGPAGPAGPAGPAG 180
QY 181 AAGPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 240
DB 181 AAGPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 240
QY 241 EVAFSHYOKAALFPDGRHRLDKIEDERLHEDVFMFTIQCNNGE 286
DB 241 EVAFSHYOKAALFPDGRHRLDKIEDERLHEDVFMFTIQCNNGE 286

RESULT 3
TNFS_HUMAN STANDARD; PRT; 261 AA.
AC P29965;

[illegible]

```

RESULT 4
PSPD_HUMAN
ID PSPD_HUMAN STANDARD; PRT: 375 AA.
AC P35247;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 198, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN D PRECURSOR (SP-D) (PSP-D).
GN PSPD OR STPD OR PSPD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxId=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA MEDLINE=93155122; Pubmed-8428971;
RA Crouch E., Rust K., Velle R., Denis-Keller H., Grosso L.;
RA "Genomic organization of human surfactant protein D (SP-D). SP-D is
RA encoded on chromosome 10q22.2-23.1."
RA J. Biol. Chem. 268:2976-2983(1993).
RX [2]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 214-243.
RC TISSUE=Lung, and Amniotic fluid;
RA MEDLINE=92322003; Pubmed-1339284;
RA Lu J., Willis A.C., Reid K.B.M.;
RA "Purification, characterization and cDNA cloning of human lung
RA surfactant protein D."
RA Biochem. J. 284:795-802(1992).
RX [3]
RP SEQUENCE OF 60-375 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Lung;
RA MEDLINE=91378578; Pubmed-1898081;
RA Rust K., Grosso L., Zhang V., Chang D., Persson A., Longmore W.,
RA Cal G.-Z., Crouch E.;
RA "Human surfactant protein D: SP-D contains a C-type lectin
RA carbohydrate recognition domain."
RA Arch. Biochem. Biophys. 290:116-126(1991).
RX [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=99197291; Pubmed-10368295;
RX Hakansson K., Lim N.K., Hoppe H.-J., Reid K.B.M.;
RX "Crystal structure of the trimetric alpha-helical coiled-coil and the
RX three lectin domains of human lung surfactant protein D."
RX Structure 7:255-264(1999).
CC -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
CC EXTENT OTHER ALPHA-GLUCOSYL MOTETES. IT COULD PARTICIPATE IN THE
CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC -1- SUBCELLULAR LOCATION: PULMONARY SURFACTANT.
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS A C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL, L05485; AAB59450.1;
DR EMBL, L05483; AAB59450.1; JOINED.
DR EMBL, L05484; AAB59450.1; JOINED.
DR EMBL, X65018; CAA46152.1;
DR PIR, A45225; A45225.
DR PIR, S18383; S18383.

```


Matches 182; Conservative 15; Mismatches 17; Indels 1; Gaps 1;

QY 257 HRLDKIEDERNLHEDFVFKTTQRCNTGERSLSLLNCEIRKQFEGFVKDMLNKEETK 316
 DB 47 HRLDKIEDERNLHEDFVFKTTQRCNTGERSLSLLNCEIRKQFEGFVKDMLNKEETK 105

QY 317 KENSFEKQKGDOPRIAHAHYSSEASSKTSVLQMAEKGYTMSNNLVTLENGKOLAVKRO 376
 DB 106 KENSFEKQKGDOPRIAHAHYSSEASSKTSVLQMAEKGYTMSNNLVTLENGKOLAVKRO 165

QY 377 GLYYIAQVTFCSNREASSQAPFIASLCLKSPGRFERILLRAANTHSSANPCGQOSIHLG 436
 DB 166 GLYYIAQVTFCSNREASSQAPFIASLCLKSPGRFERILLRAANTHSSANPCGQOSIHLG 225

QY 437 GVEPELPGASVFVWVTDPSQVSHGCTGFTSGLLKL 471
 DB 226 GVEPELPGASVFVWVTDPSQVSHGCTGFTSGLLKL 260

RESULT 8
 ID TNF5_CANFA STANDARD: PRT: 260 AA.
 AC 097626;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE CD40 LIGAND.
 GN TNF5 OR CD40LG OR CD40L.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hosie M.H., Willett B.J.;
 RT *Adjuvant properties of canine CD40L.*;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MEDIATES B-CELL PROLIFERATION IN THE ABSENCE OF CO-STIMULUS AS WELL AS ICE PRODUCTION IN THE PRESENCE OF IL-4.
 CC INVOLVED IN IMMUNOGLOBULIN CLASS SWITCHING (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- EXTRACELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS AS AN EXTRACELLULAR SOLUBLE FORM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF086711; AAD04375.1;
 DR HSSP: P29965; IALY.
 DR InterPro: IPR000478;
 DR InterPro: IPR003263;
 DR Pfam: PF00229; TNF_1;
 DR PROSITE: PS00251; TNF_1; 1.
 DR PROSITE: PS00049; TNF_2; 1.
 KM Cytokine; Transmembrane; Glycoprotein; Signal-anchor.
 FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 23 46 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 47 260 EXTRACELLULAR (POTENTIAL).
 FT DISULFID 177 217 POTENTIAL.
 FT CARBOHYD 239 239 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 260 AA; 28688 MW; 604F69A19E98EB70 CRC64;

Query Match 36.5%; Score 907.5; DB 1; Length 260;
 Best Local Similarity 81.4%; Pred. No. 5,1e-44;
 Matches 175; Conservative 18; Mismatches 21; Indels 1; Gaps 1;

QY 257 HRLDKIEDERNLHEDFVFKTTQRCNTGERSLSLLNCEIRKQFEGFVKDMLNKEETK 316
 DB 47 HRLDKIEDERNLHEDFVFKTTQRCNTGERSLSLLNCEIRKQFEGFVKDMLNKEETK 105

QY 317 KENSFEKQKGDOPRIAHAHYSSEASSKTSVLQMAEKGYTMSNNLVTLENGKOLAVKRO 376
 DB 106 KENSFEKQKGDOPRIAHAHYSSEASSKTSVLQMAEKGYTMSNNLVTLENGKOLAVKRO 165

QY 377 GLYYIAQVTFCSNREASSQAPFIASLCLKSPGRFERILLRAANTHSSANPCGQOSIHLG 436
 DB 166 GLYYIAQVTFCSNREASSQAPFIASLCLKSPGRFERILLRAANTHSSANPCGQOSIHLG 225

QY 437 GVEPELPGASVFVWVTDPSQVSHGCTGFTSGLLKL 471
 DB 226 GVEPELPGASVFVWVTDPSQVSHGCTGFTSGLLKL 260

RESULT 9
 ID CONG_BOVIN STANDARD: PRT: 371 AA.
 AC P23805;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE CONGLUTININ PRECURSOR.
 GN CGNL.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovine; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-93213261; PubMed-846093;
 RT *Cloning and sequencing of a cDNA coding for bovine conglutinin.*;
 RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
 CC [2]
 CC SEQUENCE FROM N.A.
 RA MEDLINE-93277452; PubMed-7684896;
 RT *The cDNA cloning of conglutinin and identification of liver as a primary site of synthesis of conglutinin in members of the Bovidae.*;
 RL Biochem. J. 292:157-162(1993).
 CC [3]
 CC SEQUENCE FROM N.A.
 RA TISSUE-Liver;
 RC MEDLINE-94215917; PubMed-8163202;
 RA Llou L.S., Sasstry R., Hartshorn K.L., Lee Y.M., Okarma T.B., Tauber A.I., Sasstry K.N.;
 RT *Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization of the BC cDNA reveals strong homology to surfactant protein-D.*;
 RL Gene 141:277-281(1994).
 CC [4]
 CC SEQUENCE FROM N.A.
 RA TISSUE-Semen;
 RC MEDLINE-94267222; PubMed-8207234;
 RA Llou L.S., Sasstry R., Hartshorn K.L., Lee Y.M., Okarma T.B., Tauber A.I., Sasstry K.N.;
 RT *Bovine conglutinin gene exon structure reveals its evolutionary relationship to surfactant protein-D.*;
 RL J. Immunol. 153:173-180(1994).
 CC [5]
 CC SEQUENCE OF 21-371.
 RA MEDLINE-91131556; PubMed-1993651;
 RT *Primary structure of bovine conglutinin, a member of the C-type animal lectin family.*;
 RL J. Biol. Chem. 266:2715-2723(1991).
 RN [6]
 RP PRELIMINARY SEQUENCE OF 21-52.

RA MEDLINE-87184551; PubMed-3566740;
 RT Young N.M., Leon M.A.;
 RL "The carbohydrate specificity of concanavalin and its homology to
 RL Biochem. Biophys. Res. Commun. 143:645-651(1987).
 CC -1- FUNCTION: CALCIUM-DEPENDENT LECTIN-LIKE PROTEIN WHICH BINDS TO A
 CC COMPONENT (C3BI). IT IS CAPABLE OF BINDING THROUGH THE COMPLEMENT
 CC N-ACTIVATING LECTIN, MANNOSE, AND FUCCOSE RESIDUES.
 CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D14085; BAA03170.1; -
 DR EMBL; X71774; CA50665.1; -
 DR EMBL; L18871; AAA20126.1; -
 DR EMBL; U06860; AAB60624.1; -
 DR EMBL; U06854; AAB60624.1; JOINED.
 DR EMBL; U06855; AAB60624.1; JOINED.
 DR EMBL; U06856; AAB60624.1; JOINED.
 DR EMBL; U06857; AAB60624.1; JOINED.
 DR EMBL; U06858; AAB60624.1; JOINED.
 DR EMBL; U06859; AAB60624.1; JOINED.
 DR PIR; A29416; A29416.
 DR PIR; A23740; A23740.
 DR PIR; J0450; J0450.
 DR HSSP; P08661; TRD.
 DR InterPro; IPR000087; -
 DR InterPro; IPR001304; -
 DR Pfam; PF01391; Lectin_c.1.
 DR PROSITE; PS00059; Lectin_c.1.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 DR LECTIN; PS00041; C-TYPE LECTIN_2; 1.
 KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;
 KW Collagen; Repeat; Calcium; Signal.
 FT SIGNAL 1 20
 FT CHAIN 1 20
 FT DOMAIN 46 371
 FT MOD_RES 273 371
 FT MOD_RES 63 371
 FT MOD_RES 87 371
 FT MOD_RES 99 371
 FT MOD_RES 135 371
 FT MOD_RES 141 371
 FT MOD_RES 159 371
 FT MOD_RES 162 371
 FT MOD_RES 198 371
 FT SITE 201 203
 FT DISULFID 275 369
 FT CARBOHYD 347 361
 FT CONFLICT 173 173
 FT CONFLICT 210 210
 FT CONFLICT 218 218
 FT CONFLICT 272 272
 FT SEQUENCE 371 AA; 37994 MW; 867BB41992544B1F CRC64;
 Query Match 35.3%; Score 878; DB 1; Length 371;
 Best Local Similarity 49.5%; Pred. No. 3, 1e-42;
 Matches 180; Conservative 44; Mismatches 88; Indels 52; Gaps 6;
 QY 1 MLPTSMVLYLQPLNLGAKMSLSQSRVPTCTLVMSPTENGLPDRDRDGRGREG 60
 Db 3 LTP-LTSLVLTLPWMSLGAEMTTFQKILANACTLVMSPLSLFPHDGDGRGREG 61

QY 61 EKGDPELPQMLSLGLOGTGTPVPGKPGKNGSAGEPGRGSLGPGGLPGIPGPKRG 120
 Db 62 EKGDDSGPQPGPAGRRPQWPIPKDNDGVPGPGRGD---TGPRGPPGMPGAGRG 118
 QY 121 PSKGNGNIGPGKGPQPGKPGKEVAPGMOGSTAKSGTGPGRGAPGOGAPGAG 180
 Db 119 PSKGNGSGMPGTGTGPGGTGPGGVPAGPQIQ---GFPDGLKKEKAPGTGTGAPGAG 175
 QY 181 AAGPAGPAGPQAGPAGSPPGLKGDGVPDGRGSGESLPDPAALRQOMALGKGLRL 240
 Db 176 VTGSGAIGPQSGSAGAPGLKGDGDEGTGAGESGLAEVNAIKORYTILDGHLRF 235
 QY 241 EVAFSHQKALPDPDGRRLDRTEDERNLHEDVPMKTIORCNTGERSLSLNCSEELK 300
 Db 236 QNMFSGYKKAIVLPDQANGKI-----FKTAGAVKYSDAEDL-----CREAKG- 280
 QY 301 FEGFVADIMLNKEETKRENSFEMQKGDQNPQIAHVSLSKTSYLAQKGYTMSN 360
 Db 281 -----QLASPRSAENEAVTQNVRAQERNAYLSMN 310
 QY 361 NLVT 364
 Db 311 DIST 314
 RESULT 10
 ID TNPS_MOUSE STANDARD; PRT; 260 AA.
 AC P27518;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CD40 LIGAND (TNF-RELATED ACTIVATION PROTEIN) (TRAP) (T CELL ANTIGEN
 DE GP39).
 GN TNPSF OR CD40LG OR CD40L.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92244364; PubMed-1374165;
 RA Armitage R., Fanslow W., Sato T.A., Clifford K.N., Strockbine L.,
 RA Macculfi B.M., Anderson D.M., Gimpel S.D., Davis-Smith T.,
 RA Maliszewski C.R., Clark E.A., Smith C.A., Grabstein K.H., Cosman D.,
 RA Spriggs M.K.;
 RT "Molecular and biological characterization of a murine ligand for
 RT CD40.";
 RL Nature 357:80-82(1992).
 RN [2]
 RP SIMILARITY TO THE TNF FAMILY.
 RX MEDLINE-92310561; PubMed-1377364;
 RA Farrah T., Smith C.A.;
 RT "Emerging cytokine family.";
 RL Nature 358:26-26(1992).
 RN [3]
 RP 3D-STRUCTURE MODELING OF 115-260.
 RX MEDLINE-93200072; PubMed-8095800;
 RA Pelisch M.C., Jongeneel C.V.;
 RT "A 3-D model for the CD40 ligand predicts that it is a compact trimer
 RT similar to the tumor necrosis factors.";
 RL Int. Immunol. 5:233-238(1993).
 CC -1- FUNCTION: MEDIATES B-CELL PROLIFERATION IN THE ABSENCE OF CO-
 CC STIMULUS AS WELL AS IS PRODUCTION IN THE PRESENCE OF IL-4.
 CC INVOLVED IN IMMUNOGLOBULIN CLASS SWITCHING.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS AS AN
 CC EXTRACELLULAR SOLUBLE FORM.
 CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED ON ACTIVATED CD4+
 CC T-LYMPHOCYTES.
 CC -1- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X65453; CAA6448.1; -
 DR PIR: S21738; S21738.
 DR PDB: 1CDA; 31-OCT-93.
 DR MGD: MGI:88337; Tnfsf5.
 DR InterPro: IPR000478; -
 DR Pfam: PF00229; Tnfr_1.
 DR PROSITE: PS00251; Tnfr_1; 1.
 DR PROSITE: PS50049; Tnfr_2; 1.
 DR Cytokine; Transmembrane; Glycoprotein; Signal-anchor; 3D-structure.
 KM CYTOKINE; Transmembrane; Glycoprotein; Signal-anchor (POTENTIAL).
 FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 23 46 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 (POTENTIAL).
 FT DOMAIN 47 260 EXTRACELLULAR (POTENTIAL).
 FT DISULFID 177 217 POTENTIAL.
 FT CARBOHYD 239 239 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 260 AA; 29396 MW; 7E0E34F7473668B7 CRC64;

Query Match 33.3%; Score 827; DB 1; Length 260;
 Best Local Similarity 64.8%; Pred. No. 1,4e-35;
 Matches 171; Conservative 29; Mismatches 52; Indels 12; Gaps 4;

QY 209 PGDGIKSGESGLPDSALRQOMALKGLRL-EVASHYKALFPDGHRLKIEDER 267
 DB 8 PSRPSV--ATGLPSMKIFMYLTVFLITQMGISVLAFLVY-----HRLDKVEEV 57
 QY 268 NLHEDFVPMKTIQRCNTGERSLSLNCETKSGFEGYKIDIMLKETKENSEMOKD 327
 DB 58 NLHEDFVPMKTIQRCNTGERSLSLNCETKSGFEGYKIDIMLKETKENSEMOKD 116
 QY 328 ONPDIAHVISEASKTTSVLAQAEKGYTTSNNLVLTLENKOLTVRCGLYYTAAVTF 387
 DB 117 EDPOIAAHVSEASKTTSVLAQAEKGYTTSNNLVLTLENKOLTVRCGLYYTAAVTF 176
 QY 388 CSNREASSQAFIASCILSPGFERILIRANVHSSKPCGQOSIHGLGFELQPGASV 447
 DB 177 CSNREASSQAFIASCILSPGFERILIRANVHSSKPCGQOSIHGLGFELQPGASV 236
 QY 448 FVAVTDPQSVSHGTGFTSGFLKLT 471
 DB 237 FVAVTDPQSVSHGTGFTSGFLKLT 260

RESULT 11
 CL43_BOVIN STANDARD; PRT; 301 AA.
 ID CL43_BOVIN STANDARD; PRT; 301 AA.
 AC P42916;
 DT 01-NOV-1995 (Rel. 32, Last Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE COLLECTIN-43 (CL-43).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN 11
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver.
 RX MEDLINE=94216283; PubMed=8163480.
 RA Linn B.-L., Willis A.C., Reid K.B.M., Lu J., Laursen S.B.,
 RA Jensenius J.C., Holmskov U.;
 RA "Primary structure of bovine collectin-43 (CL-43). Comparison with
 RA conglutinin and lung surfactant protein-D.";
 RT J. Biol. Chem. 269:11820-11824(1994).

CC -1- FUNCTION: LECTIN THAT BINDS TO VARIOUS SUGARS: MANNOSE - MANNAC >
 CC FUCOSE > GLCNAC > GLUCOSE - MALTOSE > GALACTOSE > LACTOSE >
 CC GALNAC. COULD PLAY A ROLE IN IMMUNE DEFENSE.
 CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- SIMILARITY: CONTAINS A COLLAGEOUS DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X75912; CAA5351.1; ALT_S80.
 DR HSSP: P19999; 1MSB.
 DR InterPro: IPR000087; -
 DR InterPro: IPR001304; -
 DR Pfam: PF01391; Collagen; 2.
 DR Pfam: PF00059; Lectin_C; 1.
 DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
 KM Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Collagen;
 KW Repeat; Calcium.
 FT DOMAIN 29 142 COLLAGEN-LIKE.
 FT DOMAIN 202 301 C-TYPE LECTIN (SHORT FORM).
 FT DISULFID 204 299 BY SIMILARITY.
 FT DISULFID 277 291 BY SIMILARITY.
 SQ SEQUENCE 301 AA; 31362 MW; 0385C10B8424CD76 CRC64;

Query Match 22.7%; Score 565.5; DB 1; Length 301;
 Best Local Similarity 35.8%; Pred. No. 4.8e-25;
 Matches 127; Conservative 42; Mismatches 81; Indels 105; Gaps 8;

QY 21 EMKLSQSRVSNPTCTLWCSPTENGILPGRDGRGRGKPGGLGPGSLGSGPT 80
 DB 2 EMDVSEKTLTDPCTLVACAPPADSLKRGDRDKEGQGGKGP----- 46
 QY 81 GPVGPCKGSGACBPCKRGESGLSPGLPGIPGPAKESGSGKGNIGPGKPGKGEA 140
 DB 47 -----GPGGM-----PGPAGREGSGSGSGSGMPGTGPGGEP 79
 QY 141 GPKEVGAPGQSGTGAKSGTGPKEGKAGVQCAPNAGAGAPGAPGAPGSGRCP 200
 DB 80 GPEGVGAPGMPGS---PGPAGLGERGAP-----GPGGAIQPGGPGGAMP 124
 QY 201 GLKGRGVPGDGRGKIGSGGLPDSALRQOMALKGLRL-EVASHYKALFPDGHRL 260
 DB 125 GLKGRGVPGDGRGKIGSGGLPDSALRQOMALKGLRL-EVASHYKALFPDGHRL 184
 QY 261 DKIEDERNLHEDFVPMKTIQRCNTGERSLSLNCETKSGFEGYKIDIMLKET 304
 DB 185 EKI-----FTKAGVAKYSDAEOL-----CREAKGQLASPRSSAENAVQIVRAK 230
 QY 305 VKDIMLKETKENSEMOKG-----DONPOIAHVISEAS 341
 DB 231 NKHAIVLSMNDISKEGKFTYPTGSGLDYSNMAPGEPGNARAKDEGPNCLEIYSDN 285

RESULT 12
 CA13_BOVIN STANDARD; PRT; 1049 AA.
 ID CA13_BOVIN STANDARD; PRT; 1049 AA.
 AC P04258;
 DT 20-MAR-1987 (Rel. 04, Last Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE COLLAGEN ALPHA 1(IIII) CHAIN.
 GN COL3A1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OX Bovidae; Bovinae; Bos.
 NCBI_TaxID=9913;
 [1]
 RP SEQUENCE OF 1-242.
 RX MEDLINE=80026026; PubMed=488906;
 RA Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wachter E.,
 Kuhn K.;
 "The covalent structure of calf skin type III collagen. I. The amino
 acid sequence of the amino terminal region of the alpha 1(III) chain
 (positions 1-222).";
 Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
 RN [2]
 RP SEQUENCE OF 243-422.
 RX MEDLINE=80026027; PubMed=488907;
 RA Dewes H., Fietzek P.P., Kuhn K.;
 "The covalent structure of calf skin type III collagen. II. The amino
 acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
 (positions 223-402).";
 Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
 RN [3]
 RP SEQUENCE OF 423-571.
 RX MEDLINE=80026028; PubMed=488908;
 RA Bentz H., Fietzek P.P., Kuhn K.;
 "The covalent structure of calf skin type III collagen. III. The
 amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
 (positions 403-551).";
 Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
 RN [4]
 RP SEQUENCE OF 572-808.
 RX MEDLINE=80026029; PubMed=488909;
 RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
 "The covalent structure of calf skin type III collagen. IV. The amino
 acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
 (positions 552-788).";
 Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
 RN [5]
 RP SEQUENCE OF 809-947.
 RX MEDLINE=80026030; PubMed=488910;
 RA Dewes H., Fietzek P.P., Kuhn K.;
 "The covalent structure of calf skin type III collagen. V. The amino
 acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
 (position 789-927).";
 Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
 RN [6]
 RP SEQUENCE OF 948-1049.
 RX MEDLINE=80026031; PubMed=488911;
 RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
 "The covalent structure of calf skin type III collagen. VI. The amino
 acid sequence of the carboxyterminal cyanogen bromide peptide alpha
 1(III)CB9B (positions 928-1028).";
 Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
 RN [7]
 RP FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 ALONG WITH TYPE I COLLAGEN.
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
 LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
 ALSO CROSS-LINKED VIA HYDROXYLINES.
 CC -1- P-TM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC PIR: A03862; CSBOTS.
 DR InterPro: IPR000087;
 DR Pfam: PF01391; Collagen. 17.
 DR PROSITE: PS01208; VWFC; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Glycoprotein; Collagen.
 FT DOMAIN 1 14
 FT MOD_RES 15 1040 NONHELICAL REGION (N-TERMINAL).
 FT MOD_RES 1041 1049 TRIPLE-HELICAL REGION.
 FT MOD_RES 95 95 NONHELICAL REGION (C-TERMINAL).
 FT MOD_RES 107 107 HYDROXYLATION.
 FT MOD_RES 119 119 HYDROXYLATION.
 FT MOD_RES 938 938 HYDROXYLATION.

FT MOD_RES 950 950 HYDROXYLATION.
 FT CARBOHYD 107 107 CROSS-LINK SITE.
 FT CARBOHYD 107 107 CROSS-LINK SITE.
 FT DISULFID 1040 1040 INTERCHAIN.
 FT DISULFID 1041 1041 INTERCHAIN.
 SQ SEQUENCE 1049 AA; 93651 MW; 8EEC3D1C6EC9A3 CRC64;
 Query Match 19.58; Score 484.5; DB 1; Length 1049;
 Best Local Similarity 41.24; Pred. No. 5,2e-20;
 Matches 113; Conservative 13; Mismatches 71; Indels 77; Gaps 9;
 QY 14 PLGKGAEMKSLQSYVNTCTLWMCSPTEGCGDGDGE-----GPGKGDGDL 67
 DB 427 PKGIDGAPGRN-GERGPG-----GPGPGAGNGBTGCPPTGSGDGDGP 478
 QY 68 PGPMGLSGLOGPTGPTGPGGNSAGEPGRKGRGL-----SGPRGLRGI 112
 DB 479 GPGGGLGILPTGS-----GPGKNGRDEPCKGEAGAPGIPGKGDSDGAPGPGAGP 535
 QY 113 PGKAGKEGSGKGNIGPGKPKPKEA-----GPKGEVAGMGQS 154
 DB 536 PGKAGGAPGPGGCGKAGAPGPGPSAGTGTGCGMGRGCGPGPKDKGEPGSSGV 595
 QY 155 TGAKGSTG-----KGEKAGGV-----QGAPNAGAPGAPGAPG 190
 DB 596 DGAPKDGPRGPTGPGPGAGCGDGDGSGAGVGIAGPRGPGEREGGPPPPAPF 655
 QY 191 QGAPSRPPTGKDRGVPGDGRIGESGLPDSA 224
 DB 656 PGAPGNBEPKGERGAPGE---KGEKGPAGAA 686
 RESULT 13
 CALL_CANFA
 ID CALL_CANFA STANDARD; PRT: 1460 AA.
 AC Q9XSJ7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 GN COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OX Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
 "Sequence of normal canine COL1A1 cDNA."
 Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN).
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -1- P-TM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL: AF153062; AAD34619.1;
 DR InterPro: IPR000885;
 DR InterPro: IPR01007;
 DR Pfam: PF01410; COLFI. 1.
 DR PROSITE: PS01208; VWFC. 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

| | | | | |
|---------------------------|--------|--------------------|-------|--------------|
| Query Match | 19.48; | Score 481.5; | DB 1; | Length 1464; |
| Best Local Similarity | 41.68; | Pred. No. 1.1e-19; | | |
| Matches 112; Conservative | 0.74 | | | |

| Best Local Similarity | 41.68; | Pred. NO. 1.1e-19; | Length 1464; |
|---------------------------|------------------------------|--------------------|--------------|
| Matches 112; Conservative | 8.4e-19; <td></td> <td></td> | | |

| | | | | | | |
|----|--------------------------------|-----|--------|-----|------|----|
| 14 | PLGNLGAEMKSI,SOBSVBNCTI VQGGST | /// | Indels | 71; | Gaps | 7; |
|----|--------------------------------|-----|--------|-----|------|----|

```

      1::|::|
      2::|
      3::|
      4::|
      5::|
      6::|
      7::|
      8::|
      9::|
     10::|
     11::|
     12::|
     13::|
     14::|
     15::|
     16::|
     17::|
     18::|
     19::|
     20::|
     21::|
     22::|
     23::|
     24::|
     25::|
     26::|
     27::|
     28::|
     29::|
     30::|
     31::|
     32::|
     33::|
     34::|
     35::|
     36::|
     37::|
     38::|
     39::|
     40::|
     41::|
     42::|
     43::|
     44::|
     45::|
     46::|
     47::|
     48::|
     49::|
     50::|
     51::|
     52::|
     53::|
     54::|
     55::|
     56::|
     57::|
     58::|
     59::|
     60::|
     61::|
     62::|
     63::|
     64::|
     65::|
     66::|
     67::|
     68::|
     69::|
     70::|
     71::|
     72::|
     73::|
     74::|
     75::|
     76::|
     77::|
     78::|
     79::|
     80::|
     81::|
     82::|
     83::|
     84::|
     85::|
     86::|
     87::|
     88::|
     89::|
     90::|
     91::|
     92::|
     93::|
     94::|
     95::|
     96::|
     97::|
     98::|
     99::|
    100::|
    101::|
    102::|
    103::|
    104::|
    105::|
    106::|
    107::|
    108::|
    109::|
    110::|
    111::|
    112::|
    113::|
    114::|
    115::|
    116::|
    117::|
    118::|
    119::|
    120::|
    121::|
    122::|
    123::|
    124::|
    125::|
    126::|
    127::|
    128::|
    129::|
    130::|
    131::|
    132::|
    133::|
    134::|
    135::|
    136::|
    137::|
    138::|
    139::|
    140::|
    141::|
    142::|
    143::|
    144::|
    145::|
    146::|
    147::|
    148::|
    149::|
    150::|
    151::|
    152::|
    153::|
    154::|
    155::|
    156::|
    157::|
    158::|
    159::|
    160::|
    161::|
    162::|
    163::|
    164::|
    165::|
    166::|
    167::|
    168::|
    169::|
    170::|
    171::|
    172::|
    173::|
    174::|
    175::|
    176::|
    177::|
    178::|
    179::|
    180::|
    181::|
    182::|
    183::|
    184::|
    185::|
    186::|
    187::|
    188::|
    189::|
    190::|
    191::|
    192::|
    193::|
    194::|
    195::|
    196::|
    197::|
    198::|
    199::|
    200::|
    201::|
    202::|
    203::|
    204::|
    205::|
    206::|
    207::|
    208::|
    209::|
    210::|
    211::|
    212::|
    213::|
    214::|
    215::|
    216::|
    217::|
    218::|
    219::|
    220::|
    221::|
    222::|
    223::|
    224::|
    225::|
    226::|
    227::|
    228::|
    229::|
    230::|
    231::|
    232::|
    233::|
    234::|
    235::|
    236::|
    237::|
    238::|
    239::|
    240::|
    241::|
    242::|
    243::|
    244::|
    245::|
    246::|
    247::|
    248::|
    249::|
    250::|
    251::|
    252::|
    253::|
    254::|
    255::|
    256::|
    257::|
    258::|
    259::|
    260::|
    261::|
    262::|
    263::|
    264::|
    265::|
    266::|
    267::|
    268::|
    269::|
    270::|
    271::|
    272::|
    273::|
    274::|
    275::|
    276::|
    277::|
    278::|
    279::|
    280::|
    281::|
    282::|
    283::|
    284::|
    285::|
    286::|
    287::|
    288::|
    289::|
    290::|
    291::|
    292::|
    293::|
    294::|
    295::|
    296::|
    297::|
    298::|
    299::|
    300::|
    301::|
    302::|
    303::|
    304::|
    305::|
    306::|
    307::|
    308::|
    309::|
    310::|
    311::|
    312::|
    313::|
    314::|
    315::|
    316::|
    317::|
    318::|
    319::|
    320::|
    321::|
    322::|
    323::|
    324::|
    325::|
    326::|
    327::|
    328::|
    329::|
    330::|
    331::|
    332::|
    333::|
    334::|
    335::|
    336::|
    337::|
    338::|
    339::|
    340::|
    341::|
    342::|
    343::|
    344::|
    345::|
    346::|
    347::|
    348::|
    349::|
    350::|
    351::|
    352::|
    353::|
    354::|
    355::|
    356::|
    357::|
    358::|
    359::|
    360::|
    361::|
    362::|
    363::|
    364::|
    365::|
    366::|
    367::|
    368::|
    369::|
    370::|
    371::|
    372::|
    373::|
    374::|
    375::|
    376::|
    377::|
    378::|
    379::|
    380::|
    381::|
    382::|
    383::|
    384::|
    385::|
    386::|
    387::|
    388::|
    389::|
    390::|
    391::|
    392::|
    393::|
    394::|
    395::|
    396::|
    397::|
    398::|
    399::|
    400::|
    401::|
    402::|
    403::|
    404::|
    405::|
    406::|
    407::|
    408::|
    409::|
    410::|
    411::|
    412::|
    413::|
    414::|
    415::|
    416::|
    417::|
    418::|
    419::|
    420::|
    421::|
    422::|
    423::|
    424::|
    425::|
    426::|
    427::|
    428::|
    429::|
    430::|
    431::|
    432::|
    433::|
    434::|
    435::|
    436::|
    437::|
    438::|
    439::|
    440::|
    441::|
    442::|
    443::|
    444::|
    445::|
    446::|
    447::|
    448::|
    449::|
    450::|
    451::|
    452::|
    453::|
    454::|
    455::|
    456::|
    457::|
    458::|
    459::|
    460::|
    461::|
    462::|
    463::|
    464::|
    465::|
    466::|
    467::|
    468::|
    469::|
    470::|
    471::|
    472::|
    473::|
    474::|
    475::|
    476::|
    477::|
    478::|
    479::|
    480::|
    481::|
    482::|
    483::|
    484::|
    485::|
    486::|
    487::|
    488::|
    489::|
    490::|
    491::|
    492::|
    493::|
    494::|
    495::|
    496::|
    497::|
    498::|
    499::|
    500::|
    501::|
    502::|
    503::|
    504::|
    505::|
    506::|
    507::|
    508::|
    509::|
    510::|
    511::|
    512::|
    513::|
    514::|
    515::|
    516::|
    517::|
    518::|
    519::|
    520::|
    521::|
    522::|
    523::|
    524::|
    525::|
```

[illegible]

db 904 ECCTCCHDCCCTA-----KGERGLSGPPGLPGICGP----- 115

...TIRKNGEVSFPGPAGEEKSSPGADGPAGAPGTPCPQGIAGQRVGL 963

-----AGKEGPSGKQGNIGPQCKPGPKGEAGPKGE-----VGA 148

[illegible]

185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696

1024 PGRDSSPGAKGDRGETGPAGPPGAPPGAPGAPVGPAGKSGDRGETGPAGPAGPVCPAC 1000

197 RCPPGL-----KGDRCVPGDRGIKGESG 219

1084 RGPAGPQGPRGDKGTGGQGDRIKGHRG 1112

SECRET

ALL_CHICK

STANDARD; PRT; 1453 AA.

21-JUL-1986 (Rel. 01, Created)
01-OCT-1989 (Rel. 13, Created)

15-JUL-1999 (Rel. 38, Last annotation update)
COLLACEN ARSW.1.1.1

COLLAI.
CHAIN PRECURSOR.

Eukaryota; Metazoa; Chordata: Craniata

Gallus. Neognathae; Galliformes; Phasianidae; Phasianinae.

```
mcbl_txid=9031;
[1]
```

SEQUENCE OF 1-153 FROM N.A.
MEDLINE=88056315

Fliner M.H., Boedtker H., Doty P.;
"Constitutional..."

of the chicken pro alpha 1(I) collagen gene.

[2].

SEQUENCE OF 1-144 FROM N.A.
MEDLINE=88007542: PubMed=38200000

Tue Jul 3 16:26:04 2001

us-09-454-223-2.rsp

Page 14

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 3, 2001, 16:14:38 ; Search time 25.5 Seconds
(without alignments)
2443.752 Million cell updates/sec

Title: US-09-454-223-2
Perfect score: 2487
Sequence: 1 MLPFLSMVLVLPPLGNLGA.....TDPQVSHGRTSGLKL 471

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rhodent:*
12: SP_unclassified:*
13: SP_vertebrate:*
14: SP_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 1039.5 | 41.8 | 378 | 6 | Q9N1X4 |
| 2 | 836 | 33.6 | 260 | 11 | Q92V2 |
| 3 | 827 | 33.3 | 260 | 11 | Q9R254 |
| 4 | 697 | 28.0 | 335 | 6 | Q97748 |
| 5 | 499.5 | 20.1 | 890 | 5 | Q77087 |
| 6 | 488.5 | 19.6 | 888 | 13 | Q90796 |
| 7 | 488 | 19.6 | 1835 | 13 | Q91A04 |
| 8 | 482.5 | 19.4 | 1445 | 13 | Q93251 |
| 9 | 481 | 19.3 | 1431 | 11 | Q93M4 |
| 10 | 480.5 | 19.3 | 1451 | 4 | Q76045 |
| 11 | 480.5 | 19.3 | 1858 | 11 | Q88207 |
| 12 | 480 | 19.3 | 1767 | 4 | Q90174 |
| 13 | 480 | 19.3 | 1818 | 4 | Q90175 |
| 14 | 480 | 19.3 | 1818 | 4 | Q90176 |
| 15 | 479.5 | 19.3 | 1838 | 4 | Q15094 |
| 16 | 479.5 | 19.3 | 1840 | 11 | Q60467 |
| 17 | 479.5 | 19.3 | 1840 | 11 | Q90103 |
| 18 | 478.5 | 19.2 | 1453 | 11 | Q63079 |
| 19 | 478 | 19.2 | 1736 | 4 | Q901P9 |

| | | | | | | |
|----|-------|------|------|----|--------|---------------------|
| 20 | 477.5 | 19.2 | 1450 | 13 | Q9Y1B4 | Q9Y1B4 cynops pyr |
| 21 | 476 | 19.1 | 284 | 4 | Q9UC14 | Q9UC14 homo sapien |
| 22 | 476 | 19.1 | 349 | 4 | Q99018 | Q99018 homo sapien |
| 23 | 476 | 19.1 | 1497 | 4 | Q9NCR9 | Q9NCR9 homo sapien |
| 24 | 476 | 19.1 | 1532 | 4 | Q02802 | Q02802 homo sapien |
| 25 | 476 | 19.1 | 1739 | 11 | Q9JL12 | Q9JL12 mus musculus |
| 26 | 475.5 | 19.1 | 1136 | 11 | Q35053 | Q35053 mus musculus |
| 27 | 475.5 | 19.1 | 1418 | 13 | Q9W7R9 | Q9W7R9 cynops pyr |
| 28 | 475 | 19.1 | 1160 | 4 | Q14046 | Q14046 homo sapien |
| 29 | 475 | 19.1 | 1433 | 11 | Q07563 | Q07563 mus musculus |
| 30 | 475 | 19.1 | 1497 | 4 | Q9UMD9 | Q9UMD9 homo sapien |
| 31 | 473 | 19.0 | 1491 | 13 | Q91718 | Q91718 xenopus lae |
| 32 | 472 | 19.0 | 1418 | 6 | Q28396 | Q28396 equus caball |
| 33 | 472 | 19.0 | 1487 | 4 | Q14047 | Q14047 homo sapien |
| 34 | 472 | 19.0 | 1487 | 6 | Q77753 | Q77753 canis famill |
| 35 | 472 | 19.0 | 1737 | 11 | Q9J104 | Q9J104 rattus norv |
| 36 | 471.5 | 19.0 | 1447 | 13 | Q91B91 | Q91B91 xenopus lae |
| 37 | 471 | 18.9 | 1486 | 13 | Q91717 | Q91717 xenopus lae |
| 38 | 470.5 | 18.9 | 473 | 11 | Q70605 | Q70605 rattus norv |
| 39 | 470 | 18.9 | 918 | 13 | Q90583 | Q90583 gallus gall |
| 40 | 469.5 | 18.9 | 1442 | 11 | Q62031 | Q62031 mus musculus |
| 41 | 469.5 | 18.9 | 1442 | 11 | Q62033 | Q62033 mus musculus |
| 42 | 469.5 | 18.9 | 1459 | 11 | Q62032 | Q62032 mus musculus |
| 43 | 469 | 18.9 | 886 | 13 | Q92029 | Q92029 gallus gall |
| 44 | 468.5 | 18.8 | 1745 | 4 | Q9NZ06 | Q9NZ06 homo sapien |
| 45 | 467.5 | 18.8 | 1142 | 4 | Q13676 | Q13676 homo sapien |

ALIGNMENTS

RESULT 1
ID : Q9N1X4 PRELIMINARY; PRT: 378 AA.
AC Q9N1X4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE LUNG SURFACTANT PROTEIN D PRECURSOR.
GN SFTPD.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20109098; PubMed-10640760;
RA van Eljk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
Lawson P.R.;
RT Porcine lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
localisation and tissue distribution.
RL J. Immunol. 164:1442-1450(2000).
DR EMBL: AF132496; AAF22145.2; -;
DR InterPro: IPR000087; -;
DR Pfam: PF00059; Lectin_C; 1.
DR Pfam: PF01391; Collagen; 3.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
DR SMART: SM00034; CLECT; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 378 LUNG SURFACTANT PROTEIN D.
SQ SEQUENCE 378 AA: 37986 MW: 350486C1E56C341D CRC64;

Query Match 41.8%; Score 1039.5; DB 6; Length 378;

Best Local Similarity 55.8%; Pred. No. 3.5e-73;

Matches 218; Conservative 36; Mismatches 82; Indels 53; Gaps 8;
QY 1 MLPFLSMVLVLPPLGNLGAEMKSLQSRSVPNCTLVMSPTENGDPGRDGRDREGPRG 60
DB 3 LLP-LSVLILLTOPPSRLGAEKMTYSQRAVANACALVMSPEKNGLGRDGRDREGPRG 61

```

QY 61 EKGDPLPGPMGLSGLOGPTGVPKGENSGABGPCKGKNGLSGPPGLPIGPBAGKEG 120
DB 62 EKDDPLPGAVRACMPGLAGVGVKGDNGSTGEGGAGDGPBGPPGPPIGPBAGKEG 121
QY 121 PSKOGNIGPOPKPGPKGAGKGVGAGGAGSTGAGSTGPKGEGAGVAGVAGPAG 180
DB 122 PSKOGNIGPOPKPGPKGAGKGVGAGGAGSTGAGSTGPKGEGAGVAGVAGPAG 181
QY 181 AAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 240
DB 182 AAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 241
QY 241 EYAFSHYOKALFPDGHRRDLKIEDERNLHEDFVFKTIQ-----RCNT 284
DB 242 OKAPSOYKAVELFPNGGVGKERI-----FKTGFEKTQDDAQVCTGAGGOMASPRSET 295
QY 285 GERSLSLNCERIKSOFEGFKDMLNKEETKENSEFMOKGDONFOIAHVISASAKT 344
DB 286 ENEALSOLVTAONKAAP-----LSMDIKTEGNFTYPTGE--PLYAN----- 336
QY 345 TSVLQMA-----EKGYTMSNNLYTL-ENGK 369
DB 337 -----MARGEPNNNGSSGAEKCEIFPNKG 362

```

```

RESULT 2
ID 0922V2 PRELIMINARY; PRT: 260 AA.
AC 0922V2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
OS CD40 LIGAND.
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN NCBI_Taxid=10116;
RP SEQUENCE FROM N.A.
RC STRAIN=PVG; TISSUE=SPLEEN.
RA Daniel K.C., Foss Y., Mousavi A., Macary P., Kemeny D.M.,
RT Farzaneh F., Gaken J.A.;
RL "Cloning and sequencing of rat CD40 ligand."
DR EMBL: AF013985; AD09323.1; -
DR HSSP: P27548; ICDA.
DR InterPro: IPR000478; -
DR Pfam: PF00229; TNE_1; -
DR PROSITE: PS008600; -
DR PROSITE: PS00251; TNE_1; 1.
DR PROSITE: PS50049; TNE_2; 1.
DR SMART: SM00207; TNE_1; 1.
SO SMART: SM00207; TNE_1; 1.
SEQUENCE 260 AA: 29282 MW; 15D21F5200FDEBB CRC64;

```

```

Query Match 33.6%; Score 836; DB 11; Length 260;
Best Local Similarity 65.5%; Pred. No. 1.5e-57;
Matches 173; Conservative 27; Mismatches 52; Indels 12; Gaps 4;
QY 209 PGRGKIGESGLPDSALRQOMALKGLQRL-EVAFSHYOKALFPDGHRRDLKIEDER 267
DB 8 PPSRSV--ATGLPASKRMIFMLTVFLITQMGISVLAAYL-----HRRLDKVEEA 57
QY 268 NLHEDFVFKTIQRCNTGERSLSLNCERIKSOFEGFKDMLNKEETKENSEFMOKGD 327
DB 58 SLHEDFVFKTKLRCKNGKSGSLNCEMRROFEDLVKDISLNKEE-KKESFEMQKGD 116
QY 328 ONFOIAHVISASAKTTSVLQMAEKGYTMSNNLYTLNGKQLTVRQGLYYIAQVTF 387
DB 117 EDPQIAAHVISEANSNAASVLAQMAKGYTMSNNLYTLNGKQLTVRQGLYYIAQVTF 176

```

```

QY 388 CSNREASSQAFPIASLCLKSPGERILLRAANTHSSAKPGGOSIHLGVFELQAGASY 447
DB 177 CSNREPLSORPFIIVSIMLKPSGSEKILLRAANTHSSSKICEQOSIHLGVFELQAGASY 236
QY 448 FVNTDPSOVSHGCFPSFGLTKL 471
DB 237 FVNTDPSOVYHIGIFSSFGLTKL 260

```

```

RESULT 3
ID 09R254 PRELIMINARY; PRT: 260 AA.
AC 09R254;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
OS CD40 LIGAND.
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN NCBI_Taxid=10116;
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY;
RA Halliott K.M., Oaks M.K.;
RT "Nucleotide sequence of the Rat CD40 ligand."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF116582; AA022460.1; -
DR HSSP: P27548; ICDA.
DR InterPro: IPR000478; -
DR Pfam: PF00229; TNE_1; -
DR PROSITE: PS00251; TNE_1; 1.
DR PROSITE: PS50049; TNE_2; 1.
DR SMART: SM00207; TNE_1; 1.
SO SMART: SM00207; TNE_1; 1.
SEQUENCE 260 AA: 29259 MW; B3D3757DE60DB73A CRC64;

```

```

Query Match 33.3%; Score 827; DB 11; Length 260;
Best Local Similarity 64.8%; Pred. No. 7.7e-57;
Matches 171; Conservative 28; Mismatches 53; Indels 12; Gaps 4;
QY 209 PGRGKIGESGLPDSALRQOMALKGLQRL-EVAFSHYOKALFPDGHRRDLKIEDER 267
DB 8 PPSRSV--ATGLPASKRMIFMLTVFLITQMGISVLAAYL-----HRRLDKVEEA 57
QY 268 NLHEDFVFKTIQRCNTGERSLSLNCERIKSOFEGFKDMLNKEETKENSEFMOKGD 327
DB 58 SLHEDFVFKTKLRCKNGKSGSLNCEMRROFEDLVKDISLNKEE-KKESFEMQKGD 116
QY 328 ONFOIAHVISASAKTTSVLQMAEKGYTMSNNLYTLNGKQLTVRQGLYYIAQVTF 387
DB 117 EDPQIAAHVISEANSNAASVLAQMAKGYTMSNNLYTLNGKQLTVRQGLYYIAQVTF 176
QY 388 CSNREASSQAFPIASLCLKSPGERILLRAANTHSSAKPGGOSIHLGVFELQAGASY 447
DB 177 CSNREPLSORPFIIVSIMLKPSGSEKILLRAANTHSSSKICEQOSIHLGVFELQAGASY 236
QY 448 FVNTDPSOVSHGCFPSFGLTKL 471
DB 237 FVNTDPSOVYHIGIFSSFGLTKL 260

```

```

RESULT 4
ID 097748 PRELIMINARY; PRT: 335 AA.
AC 097748;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
OS CONJUTININ PRECURSOR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

```

OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=93213261; PubMed=8460993;
 RA Suzuki Y., Yin Y., Makino M., Kurimura T., Makamaya N.;
 RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
 RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=93277452; PubMed=7684896;
 RA Lu J., Lauren S., Thiel S., Jensenius J., Reid K.;
 RT "The cDNA cloning of conglutinin and identification of liver as a
 primary site of synthesis of conglutinin in members of the Bovidae.";
 RL Biochem. J. 292:157-162(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=94128104; PubMed=8297370;
 RA Kawasaki N., Itoh N., Kawasaki T.;
 RT "Gene organization and 5'-flanking region sequence of conglutinin: a
 C-type mammalian lectin containing a collagen-like domain.";
 RL Biochem. Biophys. Res. Commun. 198:597-604(1994).
 DR EMBL: D25302; BAA04983.1; JOINED.
 DR EMBL: D25296; BAA04983.1; JOINED.
 DR EMBL: D25297; BAA04983.1; JOINED.
 DR EMBL: D25299; BAA04983.1; JOINED.
 DR EMBL: D25300; BAA04983.1; JOINED.
 DR EMBL: D25301; BAA04983.1; JOINED.
 DR HSSP: P35247; 1808.
 DR InterPro: IPR000087; -
 DR InterPro: IPR001304; -
 DR Pfam: PF00059; Lectin_C; 1.
 DR Pfam: PF01391; Collagen; 2.
 DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
 DR SMART: SM00034; CLECT; 1.
 FT CONFLICT 236 236 E -> V (IN REF. 2).
 FT SEQUENCE 335 AA; 34702 MW; 39D3A30BC76C134C CRC64;
 SO
 Query Match 28.0%; Score 697; DB 6; Length 335;
 Best Local Similarity 41.8%; Pred. No. 1.4e-46;
 Matches 152; Conservative 40; Mismatches 84; Indels 88; Gaps 6;
 QY 1 M L P L S M L V L L V O P L G N I G A E M K S L S O R S V P N T C T L V M C S P T E N G L P G R D G R D G R E G P R G 60
 DB 3 L L P L S V L L L T O P W R S I G A E M T F S O R L A N A C T L V M C S P L E S G L P G H Q D G D R E C P H G 61
 QY 61 E K G P G L P G P K G L S G L O G P T P V P G P K G E N S A G E R P G P G E R G L S G P P G L P G I P P G A K E G 120
 DB 62 E K G P G S P G P A G R A G R P G M V P R I G P K G D N G F V G E ----- 95
 QY 121 P S G K G N T G P G K G P K G E V G A P G M G S T G A K S T G P K G E R G A P G V O G A P G N A G 180
 DB 96 ----- P E P K D T P R G G V G A P G I Q ----- G P P G S G L G E K A P E T G A P G A G 139
 QY 181 A A G P A G P A G P G A G P P G L K D R G V P D R G I K G E S G L P D S A A L R Q O M A L K G L O R L 240
 DB 140 V T G S G A G C G P G S G A R P P L K D R D R G P E T G A K G E S G L A E V N A L K R V I L D G H L R F 199
 QY 241 E V A F S H V O K A L P P D G H R R L D K I E D E R N L H E D V F P M K T I O R C N T G E R S L S I L N C E T I K S Q 300
 DB 200 O N A S O Y K K A V L P P D G Q V A G E K I ----- F T A G A V K S Y S A E Q L ----- C R E A K G - 244
 QY 301 F E G F V K D I M L K E T K K N S F E M Q K G D N P O I A A H V I S E A S K T S Y O M A E K G Y T T M S N 360
 DB 245 ----- O L A S P R S A E N E A V T O M V R A O E K N A Y I S M N 274
 QY 361 N L V T 364
 : : |

DB 275 DIST 278
 RESULT 5
 ID 077087 PRELIMINARY; PRT; 890 AA.
 AC 077087;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE FIBRILLAR COLLAGEN CHAIN FAP1 ALPHA.
 OS Alvinella pompejana.
 OC Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Canalipalpata;
 OC Terebellida; Alvinellidae; Alvinella.
 OX NCBI_TaxID=6376;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sloc F., X., Gail F., Exposito J., Y., Garrone R., Deutsch J.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF053538; AAC35289.2; -
 DR InterPro: IPR000087; -
 DR InterPro: IPR000885; -
 DR Pfam: PF01391; Collagen; 10.
 DR Pfam: PF01410; COLFI; 1.
 DR PRODOM: PD002078; -; 1.
 DR SMART: SM00038; COLFI; 1.
 KW Collagen.
 SO SEQUENCE 890 AA; 88136 MW; 52CC756FA70CA90C CRC64;
 Query Match 20.1%; Score 499.5; DB 5; Length 890;
 Best Local Similarity 47.5%; Pred. No. 1.1e-30;
 Matches 115; Conservative 11; Mismatches 73; Indels 43; Gaps 8;
 QY 14 P L G N I G ----- A E M S L S O R S V P N T C T L V M C S P ----- T E N G L P G R D G R D G R E G P R G 63
 DB 320 P A G M T G P G P G P A G I K G L - R G E P G ----- K P R A G D G T G T P G M D G T K G E R G E D G 370
 QY 64 D P G L P G P K G L S G L O ----- G P T G P P G P K G E N S A G E P ----- S P K E R G L S G 105
 DB 371 P P G L P G P T P G P G P G E R G E V G L P E R G E P P G A P G P G E P G A P G E P D T G P R G E R S P G 430
 QY 106 P P G L P G I P G P A G K E G P S G K G N I G P O G K P ----- G P K G E A G P K G E V G A P G M G S T G A K S T G 162
 DB 431 P S G A P G A G A A G P E G P R K D L P G L A G R P D K G P P G P G L A G A P M O G L R E P P E T G 490
 QY 163 P K G E R G A P G V O G A P C N A G A P G A P A ----- G P O G A P G S R G P P G L K G D R G V P G D R G I K G E S G 219
 DB 491 P G E R G E R G E R G A G V R G P V P P P A G A P G O G S K G E R G A A G P K D K G M P G P I Q G W G 550
 QY 220 L P 221
 DB 551 T P 552
 RESULT 6
 ID 090796 PRELIMINARY; PRT; 888 AA.
 AC 090796;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE ALPHA-1 TYPE XI COLLAGEN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WHITE LEGHORN; TISSUE=LIMB;
 RX MEDLINE=93054557; PubMed=1429607;
 RA Nab H.D., Barendse M., Upholt W.B.;

QY 41 PTEMLPGRDGRRREGPRGEEKDPCLPGPMGLSGLOGPTGCVGPK-----GE 88
 Db 399 PGPPGPEPGAGQDVGGEKGEDDDPGQPPGPSPGSEAGPPGPGRCPPGATCAEGHGE 11
 QY 89 NGSAGE-----GPKGERGLSPGRLPCTPGPAPKE-----GPSKOGNITCP 130
 Db 459 XGAGGEGGAGCAGCACTGPGVPGPGAPGKPEPEELRIPGVGEQGLPGAPQDDPPGHLP 111
 QY 131 QGKRGPEGGEAGPRGEVAGAPGQSGTGAKGSTGPKGERARAVGAP-----GNACGAPAP 187
 Db 519 PGLPGLTGLKDPSGSKGEGHGLGLGLGPPGEEGEGGDLRPGSPGAKGDAITSSPAP 111
 QY 188 AGGQAGPAGSRGPPGLKGDRCVPGCDGICNGESGLP 221
 Db 579 LGPPGPPGLPGPGGPKSGKSSGSPAGQGDGSLP 612

| | | |
|----|---|---------------|
| | RESULT | 7 |
| ID | G91AUC4 | |
| AC | G91AUA4 | PRT; 1835 AA. |
| DT | 01-OCT-2000 (TEMBLrel. 15, Created) | |
| DT | 01-OCT-2000 (TEMBLrel. 15, Last sequence update) | |
| DT | 01-MAR-2001 (TEMBLrel. 16, Last annotation update) | |
| OS | ALPHA 1 (V) COLLAGEN. | |
| OS | Gallus gallus (Chicken). | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; | |
| OX | Gallus. | |
| RN | NCBI_TaxID=9031; | |
| RP | [1] | |
| RC | SEQUENCE FROM N.A. | |
| RC | STRAIN-BRED WHITE LECHORN. | |
| EX | MEDLINE=20068042; PubMed=10601735; | |
| RA | Gordon M.K., Marchant J.K., Foley J.W., Igoe F., Gibney E.P., | |
| RA | Nah H.E., Barenbaum W., Myers J.C., Rodriguez E., Dublet B., | |
| RA | van der Rest M., Linsenmayer T.F., Upholt W.B., Blk D.E.; | |
| RT | "Complete primary structure of the chicken alpha1(V) collagen chain." | |
| RL | Matrix Biol. 18:481-486(1999). | |
| DR | EMBL; AF137273; AAF8099.1.; | |
| DR | InterPro; IPRO00087.; | |
| DR | InterPro; IPRO00085.; | |
| DR | InterPro; IPRO01230.; | |
| DR | InterPro; IPRO01791.; | |
| DR | InterPro; IPRO03129.; | |
| DR | Pfam; PF01410; COLT1.1. | |
| DR | Pfam; PF02210; TSPN.1. | |
| DR | PROSITE; PS00294; PRENYLTATION; UNKNOWN_1. | |
| DR | SMART; SM00038; COLF1.1. | |
| QW | COLLAGEN. | |
| QW | SEQUENCE. 1835 AA. 18435 ME. DQF003140030.1. | |

| | | | | |
|-----------------------|-----------------|-------------------|-----------|-------------|
| Query Match | 19.6% | Score 488 | DB 13 | Length 1835 |
| Best Local Similarity | 47.5% | Pred. NO. 2.1e-29 | | |
| Matches 103 | Conservative 14 | Mismatches 64 | Indels 36 | Gaps 5 |

QY 41 PTENGLPGKDGKDGKREPRGSKGD-----PGLPGPMKLSGLGR---TGPVGPKEGNS 91
 DB 1340 PDDPGGPPGPGAGGDDGPPGDKGSDGSEPGQTGSPGPTGEGGPGSGPKGKNGPPGAPPGGR 1399
 QY 92 AGEPPRGKSGRLSGPP-----GLPLPGPKGRE-----GPSKGQGN 127
 DB 1400 GEGKAKGSEALIECPKGTGPIGQAPGKACPDGIRGIPGVGEGGGLPGSGPGPGPPGR 1459
 QY 128 IGPOGKPPPKKEADPPKREVENAPKMGOSTGAKSGTSGKSGRGAAPGVQGA---PGNAGAAP 184
 DB 1460 LGPPGLPGKLGKDSPPKKEKHPLILYLIPPGGOGKSGKRLGPGQSSAGKKGEDGTLTP 1519
 QY 185 AGPAPGPGAPGSKNPPGLKLDKRCVLPDRGIKESGSP 221
 DB 1520 SGPIGPGRPGPLGPPGPGPKGAKGSSGPTPKESGLP 1556

| | | |
|------------------|---|---|
| RESULT | B | |
| ID | 093251 | |
| AC | 093251 | PRELIMINARY; |
| DT | 01-NOV-1998 | (TREMBlrel. 08, Created) |
| DT | 01-NOV-1998 | (TREMBlrel. 08, Last sequence update) |
| DE | 01-MAR-2001 | (TREMBlrel. 16, Last annotation update) |
| OC | ALPHA 1 TYPE I COLLAGEN. | |
| OC | Rana catesbeiana (Bull frog). | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana. | |
| NCBI_TaxID=8400; | | |
| RP | SEQUENCE FROM N.A. | |
| RA | Asahina K., Utoh R., Obara M., Yoshizato K.; | |
| RT | "Spatiotemporal expression of bullfrog $\alpha 1(I)$ and $\alpha 2(I)$ collagen genes | |
| RL | in intestine during metamorphosis." | |
| DR | Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases. | |
| DR | EMBL; AB015140; BAA29028.1; | |
| DR | InterPro: IPR000087; | |
| DR | InterPro: IPR000885; | |
| DR | InterPro: IPR001007; | |
| DR | Pfam: PF01391; Collagen; 18. | |
| DR | Pfam: PF01410; COLFI; 1. | |
| DR | ProDom: PDO02078; -; 1. | |
| DR | ProSITE: PS01208; WVFC; 1. | |
| DR | SMART: SM00038; COLFI; 1. | |
| SO | SEQUENCE | 1445 AA; 137251 MW; F59BB550C9873F04 CRC64; |

| | | | | |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match | 19.48; | Score 482.5; | DB 13; | Length 1445; |
| Best Local Similarity | 44.18; | Pred. No. 4.3e-29; | | |
| Matches 108; Conservative | 11; | Mismatches 63. | Indels 23 | |

| | | | | | |
|----|------|--|------|----------------------------|-----|
| QY | 41 | PTNGLGRDRDREGREGRGEGDGLDGPMLSLGLOP | --- | TDVPKPKENGSGAGEPP | 97 |
| Db | 856 | PGSTGPPGAAGRPGPGPGSGMNGPPGSPGSPAKGBQKNGETGPADRPDEBPGLAAGPPGP | --- | | 915 |
| QY | 98 | KGER | --- | GAAGSPGLGPIGP | --- |
| Db | 916 | SGEKGSPGSDGAGAPGIPGPGGLAGTGTGVLGDRGERPPLGLPPTGPTEPKQSSSG | --- | AKGEGSPCK | 124 |
| QY | 125 | QGNIGPQKPGPKGEGAPKE | --- | VGAPGMQSGTGANGSGPPGER | --- |
| Db | 976 | SEERPPGPGSPGLAGPPGREGSPGSGSPGRGSGAGPPGDDGEGSPAPGPPAPGA | --- | GAPCV | 172 |
| QY | 173 | QGAPNA | --- | GAAGTACAGACGQAGPSRGPPGLGDR | --- |
| Db | 1036 | PGAGPPAPAPACKNDRCGETGPGAGPAGPAGARGSPGAPGARGDKGEAGEBGEGEMKCH | --- | GVGDNRKIKE | 217 |
| QY | 218 | SGLPD | 222 | | |
| Db | 1096 | RGEND | 1100 | | |


```

RESULT 9
ID Q9JMH4 PRELIMINARY: PRT: 1431 AA.
AC Q9JMH4:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE COLLAGEN TYPE XVII.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamamoto K., Inoue N., Fujimori A., Saito T., Shinkai H., Sakiyama H.;
RT "Mesocricetus auratus mRNA for type XVII collagen."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB027759; BAA94381.1;
DR InterPro: IPR000087;
DR Pfam: PF01391; Collagen; 5
SO SEQUENCE 1431 AA; 144579 MW; 4315631FEB2C9A5C CRC64;

Query Match 19.3%; Score 481; DB 11; Length 1431;
Best Local Similarity 44.3%; Pred. No. 5.5e-29;
Matches 105; Conservative 20; Mismatches 70; Indels 42; Gaps 7;

QY 14 PLGNLGAEMKSLQSRVPTNTCTLVMSPTENGRLGRDGRDREGPRGEK---GDPGLGCP 70
DB 574 PKGDMGSGPK-GDRGLPCTA-----GIRPGLGHPREGPKGKGSGIDPGMEGP 622
QY 71 MGLSGLOGPTPVPRKGENSGAPRGKRGSLGSPRLPPIPPAKGKSGKQGNIGP 130
DB 623 IGGRGLEGPMPKRGEPGPGS-GEKGGRGAGEGGPKGLPVGSGVGRGPGNSP---GP 678
QY 131 QGKGPKEGAPKGEVGAAPGQSGTGAKGSTGKGE-----RGAPGYOGA 175
DB 679 GPPGSGTGPGLRGVEVGLPVKGDKGLAGPRGPKDGEKPRGLTGEPGVRGLPGAVGE 738
QY 176 PGNAGAPGAPGAPGQ-----GAPGSRGPPGLGDRGVPGDRGKIGESGLPDS 223
DB 739 PGAGCAGMPAGPDCGCGRGEGGLTGMPGTRGLPGSPDGPVKPVGTGPGGGLPDS 795

RESULT 10
ID Q76045 PRELIMINARY: PRT: 1461 AA.
AC Q76045:
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE PRO ALPHA 1(I) COLLAGEN.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts."
RL J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=88329734; PubMed=2843432;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1)."
RL Gene 67:105-115(1988).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kulvanliemi H., Stacey A., Shikata H., Balwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen."
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE FROM N.A.
RA MEDLINE=91138770; PubMed=1995349;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins."
RL FEBS Lett. 279:9-13(1991).
RN [5]
RP SEQUENCE FROM N.A.
RA MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1)."
RL Matrix 11:375-379(1991).
RN [6]
RP SEQUENCE FROM N.A.
RA Korkko J.M., Earley J.J., Nuytlinck L., Depraeppe A., Prockop D.J.,
RA Ala-Korkko L.;
RT "Analysis of the COL1A1 and COL1A2 genes by CGE and DNA Sequencing in
RT 12 Patients with mild OI (Type I). Identification of Common Sequences
RT for Null Allele Mutations."
RL EMBL: AF017178; AAB94054.2;
DR InterPro: IPR000087;
DR InterPro: IPR000885;
DR InterPro: IPR001007;
DR Pfam: PF00093; VWC; 1.
DR Pfam: PF00391; Collagen; 18.
DR Pfam: PF01410; COLFI; 1.
DR ProDom: PD002078; 1.
DR PROSITE: PS01208; VWC; 1.
DR SMART: SM00038; COLFI; 1.
DR K collagen.
SO SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 19.3%; Score 480.5; DB 4; Length 1461;
Best Local Similarity 41.6%; Pred. No. 6.2e-29;
Matches 112; Conservative 9; Mismatches 77; Indels 71; Gaps 7;

QY 14 PLGNLGAEMKSLQSRVPTNTCTLVMSPTENGRLGRDGRDREGPRGEKDPGLGPMGL 73
DB 849 PIGNVAGAGAKGARGSGA-----PPGATGFPGAGRVGPPPSGAGCPGPGPAGK 900
QY 74 SGLGPTGVPVPRKGENSGAPRGKRGSLGSPRLPPIPPAKGKSGKQGNIGP 115
DB 901 EGGGPPGRTGPPARPREVGPGRGPRGKSGAGAGPAGAPPTPPGAGRGVGL 960
QY 116 -----AGKEGPGSKGNIGPGKDPGKGEAGPGE-----VGA 148
DB 961 PGGRGERGPGPLPGPSGEPGKQGPSGASGGERGPPGPGPLAGPPGSGREGAPAGGS 1020
QY 149 PCMGSGTGAK---GSTPKRGARVAGVAGPAGNCA-----GPPAGPAGPAGPS 196
DB 1021 PGRGSGPAGKDRGTETGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 1080
QY 197 RGPPGL-----KGRGVPGDGRGKIGESG 219
DB 1081 RGPAGPGGPRDCKKETGEGDGRGKIGHRG 1109

RESULT 11
ID O88207 PRELIMINARY: PRT: 1838 AA.

```

| | Matches | Mismatches | Indels | Gaps |
|-------------------------|-------------------|------------|--------|------|
| A) EMERALD NON-DOMINANT | 105; conservative | 10; | 66; | 57; |
| | | | | 4 |

| Variable | Indels | Gaps |
|---------------|--------|------|
| 41 Efficiency | 57 | 4 |

[illegible]

101 RGTSCBPCT DCIDDCAA
-----AGQDGFPGDKGDDGEPGQTGSPGPTGE 1382

-----EGPSGKQGNIGPQ-----GKPG----- 136

1383 P3PSGPPGKRGPPGPGAGPEGROGEKGAKEAGLEGPCKTGPICGPQGA PGKPPDGLRGI 1447

137 -----KGEAGPKGEVGCABCMOCSTCCKCCTCAGTCT

1443 PGPVGGGLPGSPGPDGPPGPMGDPT DCT KCSGCGWAGTATGTT
||::||||| | | | ||:

167 BGARCVCCG 1502

[illegible]

1503 KGLPGQSSGPKGDQGITGPSGLPPGLPGPPGPKGAKGSSGPTGPKGEAGHP 1560

13 E 13

12
SOL
UIT4

PRELIMINARY; PRT; 1767 AA.

01-MAY-2000 (Tremblay, 13. Great)

01-MAY-2000 (Tremblay, 13, last sequence update)
01-MAR-2001 (Tremblay, 13, last sequence update)

COLLAGEN TYPE XI ALPHA-1, Last annotation update)

Homo sapiens (Human)
COLIAL.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Primates; Hominidae; Homo; Homo sapiens (human).

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;

[1] GEORGENS, BROWNE

MEDLINE:20455728; PubMed-10486316.

Annunen S., Korkko J., Czarny M., Warman M.L., Brunner H G
Kaartajoen v. 1971

Crusberg J.R., Curtis M.A., Davenport C.R., Mulliken J.B., Tranebjærg L., Brooks D.G., Cox G.F.,

Kaitilla I., Krawczynski M.R., Latos-Bielenska A., Mukai S., Olesen B.P., Scharfetter S.L.H., Friedrich C.A.

Prockop D.J., Ala-Kokko L., Somer M., Viikula M., Zlotogora J.,

Splicing Mutations of 54 bp Exons in the COL1A1 Gene Cause Marshall

```

RT Syndrome but Other Mutations Cause Overlapping Marshall/Stickler
RT Phenotypes *.
Am. J. Hum. Genet. 65:974-983(1999).
DR EMBL: AF101112: AAF04724.1: -
DR EMBL: AF101079: AAF04724.1: JOINED.
DR EMBL: AF101080: AAF04724.1: JOINED.
DR EMBL: AF101081: AAF04724.1: JOINED.
DR EMBL: AF101082: AAF04724.1: JOINED.
DR EMBL: AF101083: AAF04724.1: JOINED.
DR EMBL: AF101084: AAF04724.1: JOINED.
DR EMBL: AF101085: AAF04724.1: JOINED.
DR EMBL: AF101086: AAF04724.1: JOINED.
DR EMBL: AF101087: AAF04724.1: JOINED.
DR EMBL: AF101088: AAF04724.1: JOINED.
DR EMBL: AF101089: AAF04724.1: JOINED.
DR EMBL: AF101090: AAF04724.1: JOINED.
DR EMBL: AF101091: AAF04724.1: JOINED.
DR EMBL: AF101092: AAF04724.1: JOINED.
DR EMBL: AF101093: AAF04724.1: JOINED.
DR EMBL: AF101094: AAF04724.1: JOINED.
DR EMBL: AF101095: AAF04724.1: JOINED.
DR EMBL: AF101096: AAF04724.1: JOINED.
DR EMBL: AF101097: AAF04724.1: JOINED.
DR EMBL: AF101098: AAF04724.1: JOINED.
DR EMBL: AF101099: AAF04724.1: JOINED.
DR EMBL: AF101100: AAF04724.1: JOINED.
DR EMBL: AF101101: AAF04724.1: JOINED.
DR EMBL: AF101102: AAF04724.1: JOINED.
DR EMBL: AF101103: AAF04724.1: JOINED.
DR EMBL: AF101104: AAF04724.1: JOINED.
DR EMBL: AF101105: AAF04724.1: JOINED.
DR EMBL: AF101106: AAF04724.1: JOINED.
DR EMBL: AF101107: AAF04724.1: JOINED.
DR EMBL: AF101108: AAF04724.1: JOINED.
DR EMBL: AF101109: AAF04724.1: JOINED.
DR EMBL: AF101110: AAF04724.1: JOINED.
DR EMBL: AF101111: AAF04724.1: JOINED.
DR InterPro: IPR000087: -
DR InterPro: IPR000885: -
DR InterPro: IPR001791: -
DR InterPro: IPR003129: -
DR Pfam: PF01391: Collagen, 18.
DR Pfam: PF02140: COLF1, 1.
DR Pfam: PF02210: TSPN, -; 1.
DR ProDom: PD002078: -; 1.
DR SMART: SM00038: COLF1, 1.
DR Collagen.
SQ SEQUENCE 1767 AA; 176645 MW; 72E00F4CA731B26 CRC64;

Query Match 19.3%; Score 480; DB 4; Length 1767;
Best Local Similarity 45.5%; Pred. No. 8, 5e-29;
Matches 100; Conservative 18; Mismatches 60; Indels 42; Gaps

QY 41 PTENGLPGR---DNRDREGPRGKDPGLPGMKLSLQGTG----- 82
      | : | | | | | | | | | | | | | | | | | | | |
DB 1275 PDDPPEPPELPAQGDVDVGDKEDDDPQCPGPPSGEAPPPGPKRGPPCAAGR 1334
      | : | | | | | | | | | | | | | | | | | | | |
QY 83 -----VCPKGENSAGEPQPKGERGLSGPPGLPGIPPAKGEPSR 124
      | | | | | | | | | | | | | | | | | | | |
QY 1335 QGEKGANGEGAGCPRKGTGPVOPGPAKPPPELRIRIPVVEGGLPRMAAODDPPGP 1394
      | : | | | | | | | | | | | | | | | | | | | |
DB 1395 ---MGPPGLGLKGDGSGKEGHHGLGLGPPGEGEGKGDRLPOTGSPGAKDGGI 1451
      | : | | | | | | | | | | | | | | | | | | | |
QY 182 AGPAGPAGPGAGPAGSGRPPGLKGDGVPGDGIRIGESGLP 221
      | | | | | | | | | | | | | | | | | | | |
DB 1452 GPPAGLGPAGPGLGPPGPGPKRGKNGSTGPAQKGDGSLP 1491
      | | | | | | | | | | | | | | | | | | | |
RESULT 13

```

[illegible]

| | | | | |
|-----------------------|------------------|--------------------|------------|--------------|
| Query Match | 19.3%; | Score 480; | DB 4; | Length 1818; |
| Best Local Similarity | 45.5%; | Pred. No. 8.8e-29; | | |
| Matches 100; | Conservative 18; | Mismatches 60; | Indels 43; | |

42

| | | | |
|----|------|--|------|
| Db | 1326 | PEDPPEGELPAGODGDGGGKGDSDSPQPPGPPGSGEAGPPGPPGKNGPPGMAAEGR | 1386 |
| QY | 83 | -----VPGKGENSGAGEPPGKSGRGLSGPGLPGLIPGPAKKEGPPSK | 124 |
| Db | 1386 | QGEKAKGEAGAEERPCKTGTGPPGCGAPAPGPEGLNGIPGPEDEGGLPGAAGDGGPPGPR | 1445 |
| QY | 125 | CGNIGPGSKPPKREAPKKEVGAFGMGSTGAKSTGPGSEBGVAPGVGAG--GMAA | 181 |
| Db | 1446 | ---WGPGLREYLGKDPSSKKEKCHPLILGIPPGEGEKDGRGLPGLVGSFGAKGGDGI | 1502 |
| QY | 182 | AGPAGPAGPGAGSRPPGLKGDRLVPVDDRGIKGEGSLP | 221 |
| Db | 1503 | PGPAGPLGPDPGPGPLPGQSPKGNKSTSTPAGQKDDSGLP | 1542 |

```

RESULT 15
ID 015094
AC 015094 PRELIMINARY; PRT; 1838 AA.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
GN PRO- $\alpha$ 1(I)-1 TYPE V COLLAGEN.
DS COL5A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_Taxid=9606;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92105142; PubMed=1722213;
RA Greenspan D.S.; Cheng W.; Hoffman G.G.;
RT "The pro-alpha 1(V) collagen chain. Complete primary structure,
RT distribution of expression, and comparison with the pro-alpha 1(XI)
RT collagen chain."
RL J. Biol. Chem. 266:24727-24733(1991).
RL EMBL; M76729; AAA5993.1; -.
DR InterPro: IPR000087; -.
DR InterPro: IPR000885; -.
DR InterPro: IPR001791; -.
DR InterPro: IPR003129; -.
DR Pfam: PF01339; COLLAGEN; 18.
DR Pfam: PF01410; COLF1; 1.
DR Pfam: PF02210; TSPN; 1.
DR ProDom: PD002078; -; 1.
DR SMART: SM00038; COLF1; 1.
DR SMART: SM00038; COLF1; 1.
SQ SEQUENCE 1838 AA; 183611 MW; 5078307F6E00F0BA CRC64;

```

[illegible]

```

Db      1323  PCKPKPGPDDDKGKSPGVGFPDGPFGGEGPAGQGGPGDGGDDGEPGGTGGSPGPTGE 1387
Qy      101  RKLSPPECLPGITPGPAGK-----EGPSGGKGGNGTGPQ-----GKGP----- 136
Db      1383  PGPSPGPCKRKPDPGPACBPBGHGEKAKKGAGLGGSPGGTGPAGPAGAPKPGPLGIGI 1442
Qy      137  -----KGEAGPKGEVGAAPAGMGOSTGAKRGSTGPCKE 166
Db      1443  PCPVPGEQGLPESPGEDPDPGPMGPGLGLGKIDSPGKSGKHGRLGLLGLIGPPGDEGEGKD 1502
Qy      167  RGACAGVGGA---PGMAGAAGPAGPAGGQAGPGRGPGGLKAGRGVPGDGRGIKGESGLP 221
Db      1503  RGLPGPGSSGSGPGEQGITPSSPITGPBPBPGLPGPPGPKGAKKSSGSGTGPKGGEAGHP 1560

```

Search completed: July 3, 2001, 16:17:22
Job time: 164 sec